



รายงานการวิจัยฉบับสมบูรณ์

การตรวจคะแนนจุดตัดขั้นต้นสำหรับกลุ่มสาระการเรียนรู้วิทยาศาสตร์จาก
คะแนนสอบระดับชาติขั้นพื้นฐานด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์:
การวิเคราะห์เพื่อการตัดสินใจ

Investigating Tentative Cutscores for Science Learning Area on the
Ordinary National Educational Test Scores using the Construct
Mapping Method: An Analysis for Further Judgments

วีรภัทร์ สุขศิริ
ชนม์ชกรณ วรอินทร์

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)
ปีงบประมาณ 2559



รายงานการวิจัยฉบับสมบูรณ์

การตรวจคะแนนจุดตัดขั้นต้นสำหรับกลุ่มสาระการเรียนรู้วิทยาศาสตร์จาก
คะแนนสอบระดับชาติขั้นพื้นฐานด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์:
การวิเคราะห์เพื่อการตัดสินใจ

Investigating Tentative Cutscores for Science Learning Area on the
Ordinary National Educational Test Scores using the Construct
Mapping Method: An Analysis for Further Judgments

วีรภัทร์ สุขศิริ
ชนม์ชกรณ วรอินทร์

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)
ปีงบประมาณ 2559

ชื่อโครงการ (ภาษาไทย) การตรวจคะแนนจุดตัดขั้นต้นสำหรับกลุ่มสาระการเรียนรู้วิทยาศาสตร์จากคะแนนสอบ
ระดับชาติขั้นพื้นฐานด้วยวิธีทำแผนที่ภาวะสันนิษฐาน: การวิเคราะห์เพื่อการตัดสินใจ

ประจำปีงบประมาณ 2559

จำนวนเงินที่ได้รับการสนับสนุน 100,000 บาท

ระยะเวลาทำการวิจัย 9 เดือน

ตั้งแต่ มีนาคม 2559 ถึง ธันวาคม 2559

วิรัช สุศิริ Graduate School of Education, University of California-Berkeley

ชนม์ชกร วรรณทร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม จังหวัดพิษณุโลก

บทคัดย่อ

แนวคิดและวิธีกำหนดคะแนนจุดตัดให้กับคะแนนสอบมีหลากหลายวิธี การวิจัยนี้เลือกใช้วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ เนื่องจากจุดเด่นของแนวคิดนี้คือ สารสนเทศที่ได้จากการเชื่อมโยงกระบวนการทดสอบกับกระบวนการกำหนดคะแนนจุดตัด สามารถตีความได้สอดคล้องกับวัตถุประสงค์ของการทดสอบตามโมเดลความสามารถแฝง ซึ่งการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ ตรวจสอบและเสนอแนะค่าที่สามารถกำหนดเป็นคะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากข้อสอบ ONET วิชาวิทยาศาสตร์ตั้งแต่ปีการศึกษา 2550 ถึง 2552

วิธีดำเนินการวิจัยเริ่มต้นจากการสร้างและตรวจสอบความสอดคล้องกลมกลืนระหว่างแผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์กับข้อมูลเชิงประจักษ์ จากนั้นแก้ไขแผนที่ภาวะสันนิษฐานดังกล่าวเพื่อให้ได้แผนที่ภาวะสันนิษฐานเชิงประจักษ์ ผลที่ได้นำไปตรวจสอบความสอดคล้องกลมกลืนโมเดลการวัดความสามารถทางวิทยาศาสตร์ ซึ่งจะพิจารณาเลือกโมเดลที่สอดคล้องกลมกลืนมากที่สุดไปตรวจสอบความไม่แปรเปลี่ยนของโมเดล จากนั้นจึงกำหนดคะแนนจุดตัดขั้นต้นตามโมเดลที่ไม่แปรเปลี่ยนและนำเสนอคะแนนสเกลความสามารถทางวิทยาศาสตร์แต่ละระดับความสามารถของนักเรียน ซึ่งผลวิจัยสรุปได้ดังนี้

1) แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์ของนักเรียน ประกอบด้วย 3 ระดับ คือขั้นต่ำกว่าขั้นพื้นฐาน ขั้นพื้นฐานและขั้นสูง โดยเป็นโครงสร้างที่ใช้บรรยายความสามารถนักเรียนได้อย่างเหมาะสม สามารถอธิบายลำดับความยากข้อสอบและความสามารถของนักเรียนที่ใช้ทำข้อสอบได้สอดคล้องกับแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่พัฒนาขึ้น

2) โมเดลที่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากที่สุด ในปีการศึกษา 2550 คือโมเดลพหุมิติระหว่างข้อสอบ ($\chi^2=2,652.12$, AIC=4542927.32, BIC=4543128.21) ปีการศึกษา 2551 คือโมเดลเอกมิติ ($\chi^2=2,484.80$, AIC=3894806.39, BIC=3894924.88) และปีการศึกษา 2552 คือโมเดลพหุมิติภายในข้อสอบที่ให้คะแนนแบบหลายค่า ($\chi^2=7,626.95$, AIC=3859428.41, BIC=3859643.36) และผู้วิจัยได้คัดเลือกโมเดลเพิ่มเติมอีกสามโมเดล คือโมเดลเอกมิติของปีการศึกษา 2550 โมเดลพหุมิติระหว่างข้อสอบของปีการศึกษา 2551 และโมเดลเอกมิติที่ให้คะแนนแบบหลายค่าของปีการศึกษา 2552 และเมื่อตรวจสอบความตรงข้ามกลุ่มพบว่า โมเดลทั้งหมดมีความตรงข้ามกลุ่ม สามารถให้ค่าพารามิเตอร์เชิงโครงสร้างไม่แตกต่างกัน

3) คะแนนจุดตัดขั้นต้นที่กั้นระหว่างนักเรียนชั้นต่ำกว่าขั้นพื้นฐานและขั้นพื้นฐานตั้งแต่ปีการศึกษา 2550 ถึง 2552 เท่ากับ 47, 46, และ 47 คะแนนตามลำดับ และมีคะแนนจุดตัดที่สองที่กั้นระหว่างนักเรียนชั้นพื้นฐานกับขั้นสูงเท่ากับ 57, 56, และ 59 คะแนนตามลำดับ ดังนั้นผู้วิจัยจึงกำหนดคะแนนจุดตัดขั้นต้นร่วมของแบบสอบ ONET ปีการศึกษา 2550 ถึง 2552 จากการคำนวณค่าเฉลี่ยของคะแนนจุดตัดขั้นต้นของแต่ละระดับซึ่งก็คือคะแนนสเกลเท่ากับ 47 และ 57 คะแนนตามลำดับ ซึ่งถ้านำคะแนนจุดตัดขั้นต้นเหล่านี้ไปใช้ประเมินความสามารถนักเรียนจะเกิดผลกระทบในภาพรวมทั้งสามปีการศึกษา จะพบว่านักเรียนจำนวนร้อยละ 35.09 อยู่ระดับขั้นต่ำกว่าขั้นพื้นฐาน มีนักเรียนจำนวนร้อยละ 48.64 อยู่ระดับขั้นพื้นฐาน และมีนักเรียนจำนวนร้อยละ 16.27 อยู่ระดับขั้นสูง

คำสำคัญ: คะแนนจุดตัด แผนที่ภาวะสันนิษฐาน การวัดของราล์ซ การทดสอบระดับชาติขั้นพื้นฐาน ทฤษฎีตอบสนองข้อสอบ

Title: Investigating Tentative Cutscores for Science Learning Area on the Ordinary National Educational Test Scores using the Construct Mapping Method: An Analysis for Further Judgments
 Researcher: (1) Weeraphat Suksiri, Graduate School of Education University of California-Berkeley USA,
 and (2) Chonchakorn Worain, Faculty of Education Pibulsongkram Rajabhat University, Thailand

Abstract

There have been several concepts and methods for cutscore setting. This study focuses on the Construct mapping method since its advantage in linking information obtained from test development to cutscore setting procedures. Cutscores obtained from this approach can be interpreted in terms of what students know and are able to do, corresponding to the purposes of assessment and the intended structures of what are really measured by the test, called construct maps. The main purpose of this study was to analyze, examine, and propose tentative cutscores for science for Thai sixth grade students on the Ordinary National Educational tests (ONETs) between the 2007 and 2009 academic years.

The research method began with establishing the theoretical science performance construct maps (SPCMs) for the overall dimension and the strand-based dimensions and evaluating the model qualities by fitting the student responses on the ONETs. Consequently, the results suggested the SPCMs to collapse some adjacent performance levels to get the empirical SPCMs. Series of measurement models for science performance were developed based on the ONET specifications, Thai national core curriculum for science and its relevant student performance indicators by using the Multidimensional Random Coefficients Multinomial Logit (MRCML) model framework, and evaluated the best fit model for each academic year. As the best fit models obtained, cross validations were conducted to examine model invariance properties. After the models achieved cross validity, the student performance estimates obtained from Rasch analysis using the MRCML model were transformed to scaled scores, and then the tentative cutscores were established on the transformed scales. The research results were as follows:

(1) the empirical SPCMs consisted three performance levels, Below basic, Basic, and Advanced levels. The SPCMs were able to describe appropriately student knowledge and skills on science and describe increasing orders of item difficulty corresponding to levels of student performance on the ONET items and the student's and item's locations on the Wright maps;

(2) the best fit models were the 2007 between-item multidimensional model ($\chi^2=2,652.12$, AIC=4542927.32, BIC=4543128.21), the 2008 unidimensional model ($\chi^2=2,484.80$, AIC= 3894806.39, BIC=3894924.88), and the 2009 within-item polytomous multidimensional model ($\chi^2=7,626.95$, AIC=3859428.41, BIC=3859643.36). Moreover, three more models were selected in order to complete the purposes of the study including the 2007 unidimensional model, the 2008 between-item multidimensional model, and the 2009 polytomous unidimensional model. All six models achieved cross validity properties with slightly differences in item difficulty estimates between the fitting sample and validating sample;

(3) the first tentative cutscores distinguishable between the students at the Below basic level and the Basic level were 47, 46, 47 scaled scores in the 2007, 2008, and 2009 academic years, respectively. The second tentative cutscores distinguishable between the students at the Basic level and the Advanced level were 57, 56, 59 scaled scores in the 2007, 2008, and 2009 academic years, respectively. The first and second common tentative cutscores computed from the average of its relevant ones were 47 and 57 scaled scores, respectively. The results shown that overall there were 35% of the students fell in the Below basic, 48.64 % in the Basic, and 16.27% in the Advanced level.

Key word: cutscore, construct map, Rasch measurement, ONET, item response theory

กิตติกรรมประกาศ

โครงการวิจัยนี้สำเร็จลุล่วงได้ด้วยดีเพราะได้รับความอนุเคราะห์จากบุคคลหลายฝ่าย ผู้วิจัยขอขอบพระคุณทุกท่านที่มีส่วนร่วมกับโครงการวิจัยนี้ไม่ว่าจะเป็นทุนวิจัย ข้อมูล ข้อเสนอแนะและคำแนะนำต่างๆ รวมทั้งกำลังใจและการเล็งเห็นคุณค่าของโครงการวิจัยนี้ที่มีส่วนช่วยพัฒนาการศึกษาชาติให้ดียิ่งขึ้น

ขอขอบคุณสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ที่สนับสนุนทุนอุดหนุนการวิจัยจำนวน 100,000 บาทและให้ผลการตอบข้อสอบ ONET เพื่อใช้วิเคราะห์ตามวัตถุประสงค์ของการวิจัย ขอขอบคุณเจ้าหน้าที่และนักวิชาการของ สทศ. ที่ช่วยประสานงานภายในและจัดเตรียมข้อมูล ขอขอบคุณคณะกรรมการพิจารณาข้อเสนอโครงการวิจัยที่ให้ข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่ง ส่งผลให้โครงการวิจัยมีแนวทางวิจัยที่ชัดเจนและเกิดประโยชน์ตามเจตนารมณ์ของ สทศ.

ขอขอบคุณศาสตราจารย์ Mark Wilson แห่ง University of California-Berkeley ที่ให้ความอนุเคราะห์เป็นที่ปรึกษาโครงการวิจัย ให้ข้อเสนอแนะทางวิชาการอย่างลุ่มลึกชัดเจน แนะนำเทคนิคปฏิบัติต่างๆ และตรวจสอบความถูกต้องของแนวคิดของผู้วิจัย ขอขอบคุณนักศึกษาปริญญาเอกสาขาวิชา Quantitative Methods and Evaluation และสาขาอื่นๆ รวมทั้งนักวิชาการแลกเปลี่ยนในวิชาสัมมนา Models of Assessment ที่ร่วมซักถามและให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการปรับปรุงเนื้อหาโครงการวิจัยให้รัดกุมและตรงประเด็นมากยิ่งขึ้น

ขอขอบคุณ ดร. ชุตินันท์ จันทระเสนานนท์ และนายสุพัฒน์ บัณฑิตธรรมกุล ที่ช่วยเตรียมความพร้อมให้กับโครงการวิจัยในระยะเริ่มต้นและประสานงานขอรับข้อมูลจาก สทศ. ขอขอบคุณ ดร. วิเวชา ปัญจมานนท์ ดร. วัชรภรณ์ จิตธรรมาศ อาจารย์อิทธิพงษ์ ตั้งสกุลเรืองไธ และอาจารย์ปรัชญา จันทา ที่แนะนำและติดต่อผู้เชี่ยวชาญด้านเนื้อหาวิทยาศาสตร์ให้กับผู้วิจัย

ขอขอบคุณคณะผู้เชี่ยวชาญด้านเนื้อหาวิทยาศาสตร์ที่ร่วมแลกเปลี่ยนความรู้และประสบการณ์ และให้ข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่งต่อการสร้างและปรับปรุงคำบรรยายความสามารถทางวิทยาศาสตร์ของนักเรียนและคำบรรยายคำตอบของนักเรียน และการวิเคราะห์ข้อกำหนดลักษณะแบบสอบ ONET เชิงลึก

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณบิดามารดาที่ให้กำลังใจในการดำเนินโครงการวิจัย และขออุทิศคุณประโยชน์ที่เกิดขึ้นจากการวิจัยนี้ให้กับครูอาจารย์ที่ถ่ายทอดความรู้และฝึกฝนผู้วิจัยให้มีจรรยาบรรณนักวิจัย ทักษะวิจัยและประสบการณ์วิจัยที่ดี

วีรภัทร์ สุขศิริ
ชนม์ชกรณ์ วรอินทร์

สารบัญ

	หน้า
บทที่ 1 บทนำ	
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์การวิจัย	4
ขอบเขตการวิจัย	4
นิยามศัพท์เฉพาะ	7
ประโยชน์ที่ได้รับ	9
กรอบแนวคิดของการวิจัย	11
บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	
หลักการของการกำหนดคะแนนจุดตัด	15
คุณลักษณะของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์	18
แผนที่ภาวะสันนิษฐานและคำบรรยายระดับความสามารถ	21
จุดเด่นของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์	24
แผนที่ภาวะสันนิษฐานและโมเดลการวัดของราล์ซ	27
การสร้างคำบรรยายระดับความสามารถด้วยวิธีทำแผนที่ภาวะสันนิษฐาน	34
คุณลักษณะของโมเดลเอกมิติและโมเดลพหุมิติความสามารถของนักเรียน	38
ตัวอย่างงานวิจัยที่เกี่ยวข้อง	43
บทที่ 3 วิธีดำเนินการวิจัย	
ตอนที่ 1 แนวคิดของกระบวนการวิจัย	46
เครื่องมือวิจัยและกลุ่มตัวอย่าง	46
เครื่องมือวิจัย	46
ประชากรและกลุ่มตัวอย่าง	48
การออกแบบการวิเคราะห์ข้อมูล	53
ข้อกำหนดคุณลักษณะแบบสอบ คุณสมบัติข้อสอบและ	53
ระดับความยากข้อสอบ ONET	
แผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์	57
โมเดลการวัดเอกมิติและพหุมิติ	69
การประเมินความสอดคล้องกลมกลืนของข้อสอบและโมเดล	77

สารบัญ (ต่อ)

	หน้า
ตอนที่ 2 ผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์	81
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	82
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	83
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้สารและสมบัติของสาร	83
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้แรงและการเคลื่อนที่	84
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้พลังงาน	84
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	85
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ดาราศาสตร์และอวกาศ	85
แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	86
แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์แบบภาพรวมและข้อสรุป	86
บทที่ 4 ผลการวิจัย	
ตอนที่ 1 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลการวัด	90
ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2550	90
ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2551	98
ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2552	105
ตอนที่ 2 ผลการตรวจสอบความตรงข้ามกลุ่มของโมเดลการวัด	117
ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2550	119
ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2551	125
ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2552	131
ตอนที่ 3 สเกลคะแนนความสามารถทางวิทยาศาสตร์ คะแนนจุดตัดขั้นต้นและผลกระทบ	137
สเกลคะแนนความสามารถทางวิทยาศาสตร์	138
สเกลคะแนนความสามารถในภาพรวม	138
สเกลคะแนนความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	141

สารบัญ (ต่อ)

	หน้า
สเกลคะแนนความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	142
สเกลคะแนนความสามารถสาระการเรียนรู้สารและสมบัติของสาร	143
สเกลคะแนนความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่	145
สเกลคะแนนความสามารถสาระการเรียนรู้พลังงาน	146
สเกลคะแนนความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	148
สเกลคะแนนความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ	149
สเกลคะแนนความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	150
คะแนนจุดตัดขั้นต้นและผลกระทบ	151
คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2550	153
คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2551	154
คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2552	156
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	158
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	160
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้สารและสมบัติของสาร	162
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้แรงและการเคลื่อนที่	163
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้พลังงาน	165
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	167
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้ดาราศาสตร์และอวกาศ	168
คะแนนจุดตัดขั้นต้นสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	170
บทที่ 5 สรุปผล อภิปรายผลและข้อเสนอแนะ	
สรุปผลวิจัย	172
อภิปรายผลวิจัย	174
ข้อเสนอแนะ	179
เอกสารอ้างอิง	181
ภาคผนวก	193
ประวัตินักวิจัย	230

สารบัญตาราง

ตารางที่	หน้า
3.1 ประชากรและกลุ่มตัวอย่างนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เข้าสอบ ONET วิชาวิทยาศาสตร์ ปีการศึกษา 2550 ถึง 2552	48
3.2 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2550	50
3.3 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2551	51
3.4 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2552	52
3.5 ข้อกำหนดลักษณะแบบสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6	54
3.6 โมเดลการวัดความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6	69
3.7 ความสอดคล้องของลำดับความยากข้อสอบระหว่างแผนที่ของ Wright และ ระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี	87
4.1 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2550	91
4.2 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2550	92
4.3 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2550	97
4.4 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2551	98
4.5 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2551	100
4.6 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2551	104
4.7 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2552	107
4.8 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2552	108
4.9 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2552	115
4.10 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลเอกมิติ ปีการศึกษา 2550	119
4.11 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2550	121
4.12 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลพหุมิติ ปีการศึกษา 2550	122
4.13 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติ ปีการศึกษา 2550	124
4.14 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลเอกมิติ ปีการศึกษา 2551	125
4.15 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2551	127

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.16 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลพหุมิติ ปีการศึกษา 2551	128
4.17 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติ ปีการศึกษา 2551	130
4.18 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลเอกมิติ ปีการศึกษา 2552	131
4.19 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2552	133
4.20 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจาก โมเดลพหุมิติ ปีการศึกษา 2552	134
4.21 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติ ปีการศึกษา 2552	136
4.22 ค่าสถิติพื้นฐานค่าประมาณความสามารถในภาพรวม	138
4.23 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	141
4.24 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	142
4.25 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้สารและสมบัติของสาร	143
4.26 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่	145
4.27 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้พลังงาน	146
4.28 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	148
4.29 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ	149
4.30 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์ และเทคโนโลยี	151
4.31 ระดับความยากข้อสอบ ONET ตามแผนที่ภาวะสันนิษฐานเชิงประจักษ์	152

สารบัญภาพ

ภาพที่	หน้า
1.1	กรอบแนวคิดของวิเคราะห์คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์
2.1	ความสัมพันธ์ระหว่างคะแนนจุดตัดและมาตรฐานความสามารถ
2.2	การประยุกต์กรอบการสร้างโมเดลภาวะสันนิษฐานเพื่อกำหนดคะแนนจุดตัด
2.3	ตัวอย่างคำบรรยายระดับความสามารถทางวิทยาศาสตร์และทักษะการเขียน
2.4	ตัวอย่างแผนที่ภาวะสันนิษฐานการให้เหตุผลทางวิทยาศาสตร์
2.5	ตัวอย่างแผนที่ภาวะสันนิษฐานทักษะการสื่อสาร
2.6	ตัวอย่างโค้งคุณลักษณะข้อสอบ
2.7	ฟังก์ชันการตอบสนองข้อสอบจำนวน 9 ข้อ
2.8	ความสัมพันธ์ระหว่างค่าประมาณความสามารถแฝงกับค่าพารามิเตอร์ ความยากข้อสอบ
2.9	ตัวอย่างของแผนที่ของ Wright จากโมเดลเอกมิตี
2.10	ตัวอย่างของแผนที่ของ Wright จากโมเดลพหุมิตี
2.11	เส้นสมมติความต่อเนื่องของคุณลักษณะของตัวแปรแฝง
2.12	ตัวอย่างคำบรรยายสรุประดับความเข้าใจเกี่ยวกับการคูณและการหาร
2.13	ตัวอย่างตำแหน่งของเส้นแบ่งระดับความเข้าใจของนักเรียนเกี่ยวกับการคูณ และการหาร
2.14	โมเดลการวัดแบบพหุมิตีและโมเดลการวัดแบบเอกมิตี
2.15	โครงสร้างความสามารถทางวิทยาศาสตร์ตามหลักสูตรแกนกลางฯ พ.ศ. 2551
3.1	กรอบการดำเนินการวิจัย
3.2	โครงสร้างความยากข้อสอบ ONET และความสามารถทางวิทยาศาสตร์
3.3	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต
3.4	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม
3.5	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้สารและสมบัติของสาร
3.6	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้แรงและการเคลื่อนที่
3.7	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้พลังงาน
3.8	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก
3.9	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ดาราศาสตร์และอวกาศ
3.10	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์ และเทคโนโลยี
3.11	แผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์

สารบัญญภาพ (ต่อ)

ภาพที่	หน้า
3.12 โมเดลเอกมิตีแบบรวมมิตีความสามารถทางวิทยาศาสตร์ของนักเรียน	70
3.13 โมเดลเอกมิตีแบบแยกรายมิตีความสามารถทางวิทยาศาสตร์ของนักเรียน	72
3.14 โมเดลพหุมิตีระหว่างข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียน	73
3.15 โมเดลพหุมิตีภายในข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียนที่กำหนดให้ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8	75
3.16 โมเดลพหุมิตีภายในข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียนที่กำหนดให้ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8	76
3.17 ตัวอย่างการกำหนดคะแนนจุดตัดด้วยวิธีสร้างแผนที่ตัวแปรเชิงทฤษฎี	80
3.18 ตัวอย่างการกำหนดคะแนนจุดตัดระหว่างปีการศึกษา	81
3.19 แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์	89
4.1 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2550	96
4.2 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2551	103
4.3 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2552	114
4.4 แผนที่ของ Wright ของโมเดลเอกมิตีจากการตรวจสอบความตรงข้ามกลุ่ม ปีการศึกษา 2550	120
4.5 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิตี ปีการศึกษา 2550	120
4.6 แผนที่ของ Wright ของโมเดลพหุมิตีระหว่างข้อสอบจากการตรวจสอบความตรงข้ามกลุ่ม ปีการศึกษา 2550	123
4.7 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิตีระหว่างข้อสอบ ปีการศึกษา 2550	124
4.8 แผนที่ของ Wright ของโมเดลเอกมิตีจากการตรวจสอบความตรงข้ามกลุ่ม ปีการศึกษา 2551	126
4.9 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิตี ปีการศึกษา 2551	126
4.10 แผนที่ของ Wright ของโมเดลพหุมิตีระหว่างข้อสอบจากการตรวจสอบความตรงข้ามกลุ่ม ปีการศึกษา 2551	129
4.11 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิตีระหว่างข้อสอบ ปีการศึกษา 2551	130
4.12 แผนที่ของ Wright ของโมเดลเอกมิตีจากการตรวจสอบความตรงข้ามกลุ่ม ปีการศึกษา 2552	132
4.13 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิตี ปีการศึกษา 2552	132

สารบัญญภาพ (ต่อ)

ภาพที่	หน้า
4.14 แผนทีชอง Wright ของโมเดลพหุมิติภายในข้อสอบจากการตรวจสอบ ความตรงข้ามกลุ่ม ปีการศึกษา 2552	135
4.15 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิติภายในข้อสอบ ปีการศึกษา 2552	136
4.16 สเกลคะแนนความสามารถทางวิทยาศาสตร์แบบภาพรวม	139
4.17 สเกลคะแนนความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	141
4.18 สเกลคะแนนความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	143
4.19 สเกลคะแนนความสามารถสาระการเรียนรู้สารและสมบัติของสาร	144
4.20 สเกลคะแนนความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่	146
4.21 สเกลคะแนนความสามารถสาระการเรียนรู้พลังงาน	147
4.22 สเกลคะแนนความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	148
4.23 สเกลคะแนนความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ	150
4.24 สเกลคะแนนความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	151
4.25 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2550	153
4.26 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2551	155
4.27 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2552	157
4.28 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้สิ่งมีชีวิตและกระบวนการดำรงชีวิต	159
4.29 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม	161
4.30 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้สารและสมบัติของสาร	163
4.31 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่	165
4.32 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้พลังงาน	166
4.33 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก	168
4.34 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ	170
4.35 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์ และเทคโนโลยี	171

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การกำหนดคะแนนจุดตัด (cutscore setting) ได้รับความสนใจเพิ่มมากขึ้นในปัจจุบัน เนื่องจากหน่วยงานที่เกี่ยวข้องกับการจัดการศึกษาใช้คะแนนจุดตัดเป็นตัวบ่งชี้ประสิทธิภาพและประสิทธิผลของระบบการศึกษา รวมถึงใช้แสดงความรับผิดชอบที่สามารถตรวจสอบได้ (accountability) ของหน่วยงานการพัฒนาศักยภาพนักเรียนให้เป็นไปในทิศทางที่เหมาะสม (Kane, 2001; Koretz & Hamilton, 2006; MacCann & Stanley, 2006; Madaus, Russell, & Higgins, 2009; Wyse, 2013) การกำหนดคะแนนจุดตัดเป็นกระบวนการระบุค่าเพื่อจำแนกนักเรียนเป็นกลุ่มตามระดับความสามารถจากคะแนนสอบ (Berk, 1986; Cizek, 2012; Draney & Wilson, 2011; Hambleton et al., 2000; Kane & Tannenbaum, 2013) คะแนนจุดตัดจึงจำแนกนักเรียนออกเป็นอย่างน้อยสองกลุ่มคือ กลุ่มผ่านซึ่งหมายถึงนักเรียนที่มีคะแนนอย่างน้อยที่สุดเท่ากับคะแนนจุดตัด และกลุ่มไม่ผ่านซึ่งหมายถึงนักเรียนที่มีคะแนนต่ำกว่าคะแนนจุดตัด แต่อย่างไรก็ตามมาตรฐานความสามารถโดยทั่วไปประกอบด้วยคะแนนจุดตัดหลายค่า เช่น มาตรฐานความสามารถของนักเรียนจากโปรแกรมการทดสอบทางการศึกษาขั้นพื้นฐานของรัฐต่างๆ ในประเทศสหรัฐอเมริกาที่ประกอบด้วยคะแนนจุดตัดสามค่าซึ่งแบ่งความสามารถนักเรียนเป็นสี่ระดับคือ ระดับต่ำกว่าพื้นฐาน (Below basic) ระดับพื้นฐาน (Basic) ระดับเชี่ยวชาญ (Proficient) และระดับสูง (Advanced)

แม้ว่าการนำคะแนนจุดตัดไปใช้จะก่อให้เกิดผลลัพธ์ไม่พึงประสงค์บางประการ (Hambleton, 1998; Madaus, Russell, & Higgins, 2009; Mehrens & Cizek, 2012; Zieky, 2012) เช่น หน่วยงานต้นสังกัดเข้าแทรกแซงอำนาจการบริหารงานของโรงเรียนที่มีนักเรียนจำนวนมากไม่ผ่านคะแนนจุดตัดขั้นต่ำสุด ความคลาดเคลื่อนของการตัดสินผลสัมฤทธิ์ทางการเรียนของนักเรียนที่มีคะแนนใกล้เคียงกับคะแนนจุดตัด และการลงโทษครูที่ไม่สามารถยกระดับความสามารถนักเรียนได้ แต่คะแนนจุดตัดก็ให้ประโยชน์มากมายต่อระบบการศึกษา นอกเหนือจากใช้บ่งชี้ระดับผลสัมฤทธิ์ของนักเรียน การบรรลุตามมาตรฐานการเรียนรู้หรือสาระการเรียนรู้ ประสิทธิภาพการเรียนการสอน คุณภาพหลักสูตรและประสิทธิภาพการบริหารการศึกษาตั้งแต่ระดับท้องถิ่นถึงระดับประเทศแล้ว ยังใช้ประเมินความพร้อมนักเรียนเพื่อเข้าเรียนในหลักสูตรเฉพาะ ใช้ประเมินคุณภาพนักเรียนก่อนสำเร็จการศึกษา และใช้คัดเลือกนักเรียนเพื่อศึกษาต่อในระดับที่สูงขึ้นอีกด้วย (Cizek, 1996, 2012; Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006; Katz, Tannenbaum, & Kannan, 2009; Phillips, 2010) สำหรับในประเทศไทยมีการทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Educational Testing) หรือ ONET เป็นการทดสอบผลสัมฤทธิ์ทางการเรียนตามมาตรฐานการเรียนรู้

ที่กำหนดไว้ในหลักสูตรการศึกษาชาติ ทำให้คะแนนสอบ ONET จึงถูกนำไปใช้ประเมินผลความสามารถนักเรียนและคุณภาพการศึกษา (ศิริพันธ์ ดิยะวงศ์สุวรรณ, ณัฐภรณ์ หลาวทอง, และศิริชัย กาญจนวาสี, 2556) รวมทั้งในปัจจุบันยังไม่มีกำหนดคะแนนจุดตัดจากคะแนนสอบ ONET จึงทำให้การกำหนดคะแนนจุดตัดดังกล่าวเป็นความท้าทายและเกิดประโยชน์อย่างกว้างขวางต่อระบบการศึกษาไทย

ข้อเสนอแนะของศิริพันธ์ ดิยะวงศ์สุวรรณ, ณัฐภรณ์ หลาวทอง และศิริชัย กาญจนวาสี (2556) ชี้ให้เห็นถึงวิกฤตคุณภาพการศึกษาไทยและความจำเป็นเร่งด่วนที่ต้องมีคะแนนจุดตัดเพื่อกระตุ้นให้หน่วยงานที่เกี่ยวข้องเร่งพัฒนาคุณภาพนักเรียน คะแนนสอบ ONET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ตั้งแต่ปีการศึกษา 2550 ถึง 2556 และชั้นประถมศึกษาปีที่ 6 ตั้งแต่ปีการศึกษา 2552 ถึง 2556 ยกเว้นปีการศึกษา 2554 มีค่าเฉลี่ยวิชาภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม คณิตศาสตร์ วิทยาศาสตร์และภาษาอังกฤษต่ำกว่าร้อยละ 50 (<http://www.onetresult.niets.or.th/AnnouncementWeb/Notice/FrBasicStat.aspx> เข้าถึงเมื่อวันที่ 26 ธันวาคม 2557) สะท้อนให้เห็นว่านักเรียนส่วนใหญ่มีผลการเรียนรู้ไม่บรรลุตามมาตรฐานการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (สำนักงานเลขาธิการสภาการศึกษา, 2552) นอกจากนี้ผลประเมินความสามารถของนักเรียนไทยวิชาวิทยาศาสตร์ วิชาการอ่านและวิชาคณิตศาสตร์ระดับนานาชาติ หรือ PISA (Program for International Student Assessment) ในปี 2006, 2009 และ 2012 แสดงให้เห็นว่านักเรียนไทยมีคะแนนเฉลี่ยทั้งสามวิชาต่ำกว่าค่าเฉลี่ยรวมจาก 57 ประเทศที่เข้าสอบ และเมื่อเปรียบเทียบกับประเทศอื่นๆ ในภูมิภาคเอเชียตะวันออกเฉียงใต้ พบว่าในปี 2006 และ 2009 ประเทศไทยมีค่าเฉลี่ยสูงกว่าประเทศอินโดนีเซียและกัมพูชา (OECD, 2007, 2010) และในปี 2012 มีค่าเฉลี่ยสูงกว่าประเทศมาเลเซีย อินโดนีเซียและกัมพูชา (OECD, 2014)

วิกฤตคุณภาพการศึกษาข้างต้นเป็นสาเหตุสำคัญที่ประเทศไทยจำเป็นต้องกำหนดคะแนนจุดตัดเพื่อบ่งชี้ประสิทธิภาพและประสิทธิผลของระบบการศึกษาเพราะคะแนนจุดตัดให้สารสนเทศที่เป็นประโยชน์ต่อนักเรียน ครู และผู้ที่เกี่ยวข้องกับระบบการศึกษาทุกระดับ เมื่อพิจารณาวิธีกำหนดคะแนนจุดตัดในปัจจุบันพบว่ามีหลายวิธี งานวิจัยของ Berk (1986) Kaftandjiev (2010) Hambleton, Jaeger, Plake และ Mills (2000) และ Reckase (2000) แสดงให้เห็นว่านับตั้งแต่อดีตจนถึงปัจจุบันมีวิธีกำหนดมาตรฐาน (standard setting) หรือคะแนนจุดตัดเกือบหนึ่งร้อยวิธี โดยที่แต่ละวิธีมีหลักการและขั้นตอน และให้ผลลัพธ์แตกต่างกัน แต่อย่างไรก็ตามเมื่อพิจารณาคุณสมบัติของแต่ละวิธีพบว่า วิธีกำหนดคะแนนจุดตัดจะเกี่ยวข้องกับเนื้อหาที่ส่วน (Skorupski, 2012) และเกี่ยวข้องกับตรรกะการพัฒนาคะแนนจุดตัดสี่ประการ (Wilson, 2015) กล่าวคือ การกำหนดคะแนนจุดตัดมีเนื้อหาส่วนแรกเกี่ยวกับแนวคิดของจุดแบ่งระดับความสามารถบนสเกลคะแนนสอบ ส่วนที่สองเป็นรายละเอียดของขั้นตอนกำหนดคะแนนจุดตัด ส่วนที่สามเป็นบทบาทของคณะผู้เชี่ยวชาญที่พิจารณาและเสนอแนะจุดตัดที่เหมาะสมให้กับผู้มีอำนาจตัดสินใจ และส่วนสุดท้ายเป็นผลกระทบที่เกิดขึ้นจากการนำคะแนนจุดตัดไปใช้

สำหรับตรรกะการพัฒนาคะแนนจุดตัดมีหลักการสำคัญประการแรกคือ การกำหนดคะแนนจุดตัดเป็นการกำหนดขอบเขตผลการเรียนรู้และบรรยายจุดมุ่งหมายของผลการเรียนรู้ที่เกิดขึ้น เช่น นักเรียน X ที่กำลังเรียนเนื้อหา Y ควรจะมีความสามารถตามมาตรฐานการเรียนรู้ A, B และ C หมายความว่านักเรียน X ควรจะมีผลการเรียนรู้ (รู้อะไรและไม่รู้อะไร มีทักษะอะไรและไม่มีทักษะอะไรทำอะไรได้และทำอะไรไม่ได้) ตามที่ระบุในมาตรฐานการเรียนรู้ A, B และ C ซึ่งเป็นขอบเขตของผลการเรียนรู้จากเนื้อหา Y ประการที่สอง การกำหนดคะแนนจุดตัดเป็นการตัดสินความสามารถของนักเรียนว่าควรมีเท่าใดถึงจะเพียงพอและบรรลุมาตรฐานการเรียนรู้ตามขอบเขตและจุดมุ่งหมายที่กำหนด ประการที่สาม การกำหนดคะแนนจุดตัดจะให้สารสนเทศที่ชัดเจนมากขึ้นว่า นักเรียนมีความสามารถในแต่ละมาตรฐานการเรียนรู้ดีเพียงใดเมื่อพิจารณาจากคะแนนสอบ กล่าวคือนักเรียนที่มีผลการเรียนรู้ไม่บรรลุมาตรฐานการเรียนรู้จะมีความสามารถต่ำกว่าระดับที่คาดหวังเพียงใด และนักเรียนที่มีผลการเรียนรู้บรรลุมาตรฐานการเรียนรู้จะมีความสามารถเหนือกว่าระดับที่คาดหวังเพียงใด และประการสุดท้าย การกำหนดคะแนนจุดตัดเป็นการตัดสินความสามารถของนักเรียนที่ยอมรับได้ นั่นคือเป็นการกำหนดคะแนนจุดตัดเพื่อจัดกลุ่มนักเรียนตามความสามารถหรือคะแนนสอบออกเป็นอย่างน้อยสองกลุ่ม คือกลุ่มนักเรียนที่มีความสามารถในระดับที่ยอมรับได้และกลุ่มนักเรียนที่มีความสามารถต่ำกว่าระดับที่ยอมรับได้

เมื่อพิจารณาเนื้อหาของการกำหนดคะแนนจุดตัดร่วมกับตรรกะการกำหนดคะแนนจุดตัดจะพบว่า มีข้อควรพิจารณาในการเลือกวิธีกำหนดคะแนนจุดตัดสำหรับแบบสอบ ONET สามประการ ประการแรกคือ วิธีที่ใช้ไม่ควรฝ่าฝืนธรรมชาติของแบบสอบที่วัดมากกว่าหนึ่งเนื้อหาความรู้และธรรมชาติของความสามารถที่นักเรียนใช้ทำข้อสอบซึ่งเป็นทักษะการคิดตั้งแต่ระดับพื้นฐานจนถึงระดับสูง (Bunch, 2012; Draney & Wilson, 2011; Kington & Tiemann, 2012; Wyse, 2013) ประการที่สองคือ คะแนนจุดตัดควรเชื่อมโยงอย่างชัดเจนกับข้อกำหนดลักษณะแบบสอบ (test specification) ระดับความสามารถของนักเรียน คำอธิบายระดับความสามารถ (Cizek, 2012; Wyse, 2013, 2014) และสมรรถนะที่ดี (Kane, 1994, 2001) เนื่องจากคะแนนสอบ ONET ถูกนำไปใช้บ่งชี้คุณภาพการศึกษาในภาพรวมและรายมาตรฐานการเรียนรู้หรือสาระการเรียนรู้ และประการสุดท้ายคือ คะแนนจุดตัดควรสอดคล้องกับมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา (the Standards for Educational and Psychological Testing: AERA, APA, & NCME, 2014) เพื่อให้คะแนนจุดตัดที่กำหนดขึ้นน่าเชื่อถือ ถูกต้อง และเป็นที่ยอมรับมากที่สุด

การเลือกใช้วิธีกำหนดคะแนนจุดตัดที่เหมาะสมสำหรับการสอบ ONET ยังมีองค์ประกอบที่ต้องคำนึงหลายประการ โครงการวิจัยนี้เลือกใช้วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ (Construct Mapping method: Draney & Kennedy, 2010; Draney, Kennedy, Moore, & Morell, 2010; Draney & Wilson, 2011; Wilson & Draney, 2002; Wyse, 2013, 2014) เนื่องจากวิธีดังกล่าวสามารถเชื่อมโยงกระบวนการกำหนดคะแนนจุดตัดเข้ากับกระบวนการทดสอบ นั่นคือสารสนเทศเกี่ยวกับแบบสอบ ONET จะเชื่อมต่อกับกระบวนการกำหนดคะแนนจุดตัด และเมื่อ

สารสนเทศเชื่อมต่ออย่างเป็นระบบกับคะแนนจุดตัด จะทำให้คะแนนจุดตัดที่กำหนดขึ้นสัมพันธ์กับวัตถุประสงค์การทดสอบ ONET และเป้าหมายการนำคะแนนจุดตัดไปใช้ ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจมั่นใจได้ว่าคะแนนจุดตัดนั้นมาจากข้อมูลที่มีคุณภาพสูง สามารถตีความและสรุปอ้างอิงเกี่ยวกับความสามารถของนักเรียนได้อย่างถูกต้องและแม่นยำอีกด้วย

โครงการวิจัยนี้มีหลายขั้นตอน ผลลัพธ์ที่เกิดขึ้นในแต่ละขั้นตอนจะเป็นสารสนเทศนำเข้าให้กับขั้นตอนต่อไปอย่างเป็นระบบ นำไปสู่ผลลัพธ์ปลายทางคือคะแนนจุดตัดขั้นต้น (tentative cutscore) ตรรกะการกำหนดคะแนนจุดตัดของโครงการวิจัยนี้เป็นแนวคิดใหม่ที่เชื่อมโยงกับแนวคิดการสร้างแบบสอบ ONET ที่มีข้อจำกัดหลายประการ โดยเฉพาะอย่างยิ่งการละเลยการออกแบบโมเดลความสามารถนักเรียนก่อนที่จะสร้างแบบสอบ ONET คะแนนจุดตัดขั้นต้นที่กำหนดขึ้นจะเป็นหลักฐานความตรง (validity evidence) ให้กับกระบวนการทดสอบ ONET และกระบวนการกำหนดมาตรฐานความสามารถของนักเรียนในอนาคต โครงการวิจัยนี้มุ่งตอบคำถามสองประการ คือ ประการแรก แผนที่ภาวะสันนิษฐาน (construct map) หรือโครงสร้างความสามารถแฝงที่ใช้จำแนกระดับความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากคำตอบข้อสอบ ONET มีลักษณะอย่างไร และประการที่สองคะแนนจุดตัดขั้นต้นที่สัมพันธ์กับแผนที่ภาวะสันนิษฐานมีค่าเท่าใดบ้างและสามารถตีความคะแนนจุดตัดขั้นต้นเหล่านั้นได้อย่างไร

วัตถุประสงค์การวิจัย

โครงการวิจัยนี้มีวัตถุประสงค์หลักคือเพื่อวิเคราะห์ ตรวจสอบและนำเสนอค่าที่สามารถกำหนดเป็นคะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากข้อสอบ ONET วิชาวิทยาศาสตร์ตั้งแต่ปีการศึกษา 2550 ถึง 2552 โดยมีวัตถุประสงค์ย่อยดังนี้

- 1) เพื่อสร้างและตรวจสอบความถูกต้องของแผนที่ภาวะสันนิษฐานสำหรับใช้เป็นคำบรรยายระดับความสามารถทางวิทยาศาสตร์ในกระบวนการกำหนดคะแนนจุดตัดขั้นต้น
- 2) เพื่อตรวจสอบความเป็นมิติ (dimensionality) ของแบบสอบ ONET และตรวจสอบความตรงข้ามกลุ่ม (cross validation) ของโมเดลความสามารถทางวิทยาศาสตร์
- 3) เพื่อเสนอค่าที่สามารถกำหนดเป็นคะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ของแต่ละปีการศึกษา และวิเคราะห์ผลกระทบที่อาจจะเกิดขึ้นโดยพิจารณาจากร้อยละของจำนวนนักเรียนที่มีคะแนนสอบ ONET ต่ำกว่าและสูงกว่าคะแนนจุดตัดขั้นต้น

ขอบเขตการวิจัย

ขอบเขตของโครงการวิจัยนี้มีเนื้อหาครอบคลุมทั้งธรรมชาติของคำตอบข้อสอบ ONET จุดมุ่งหมายของโครงการวิจัย และวิธีดำเนินการวิจัย ซึ่งเนื้อหาทั้งหมดจะช่วยให้ผลวิจัยถูกนำไปใช้ได้ อย่างถูกต้อง รายละเอียดมีดังต่อไปนี้

1) การวิจัยนี้เป็นการวิจัยเชิงบรรยาย (descriptive research) ที่ปรับแนวคิดการกำหนดคะแนนจุดตัดให้เหมาะสมกับบริบทการทดสอบขนาดใหญ่ในประเทศไทยที่ใช้แบบสอบ ONET เป็นเครื่องมือตรวจสอบคุณภาพการศึกษา ธรรมชาติของการกำหนดคะแนนจุดตัดทุกวิธีจะเกี่ยวข้องกับการตัดสินใจโดยผู้เชี่ยวชาญและผู้มีอำนาจโดยตรง วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์นี้จะนำคำตอบของนักเรียนที่มีต่อแบบสอบ ONET วิชาวิทยาศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2550 ถึง 2552 มาวิเคราะห์เพื่อประมาณค่าความสามารถของนักเรียน แล้วจึงเสนอแนะค่าที่เหมาะสมที่จะเป็นคะแนนจุดตัดขั้นต้น วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์จึงสร้างความเชื่อมั่นให้กับผู้มีอำนาจตัดสินใจว่าสารสนเทศที่ส่งไปถึงขั้นตอนคัดเลือกหรือแก้ไขคะแนนจุดตัดขั้นต้นให้เป็นคะแนนจุดตัดจริงนั้นมาจากข้อมูลที่ถูกต้อง มีคุณภาพสูงและครบถ้วน แต่อย่างไรก็ตามผู้มีอำนาจตัดสินใจจะไม่นำคะแนนจุดตัดขั้นต้นตามที่นักวิจัยเสนอไปใช้กำหนดเป็นคะแนนจุดตัดจริงก็ได้ (Atkinson, 2012; Madaus, Russell, & Higgins, 2009; Phillips, 2012; Shepard, 1979 อ้างใน Cizek, 2012; Skorupski, 2012) ดังนั้นขอบเขตของโครงการวิจัยนี้จึงไม่ครอบคลุมขั้นตอนพิจารณาตัดสินใจโดยผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจเพื่อให้ได้คะแนนจุดตัดจริง แต่เป็นการวิเคราะห์ข้อมูลและเชื่อมโยงสารสนเทศที่มีอยู่เข้าด้วยกัน แล้วนำเสนอคะแนนจุดตัดขั้นต้นที่เป็นไปได้เพื่อประกอบการพิจารณาในขั้นตอนต่อไป

2) แม้ว่าคะแนนสอบ ONET เฉลี่ยของนักเรียนชั้นประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 ของ 5 วิชาโดยเฉพาะวิชาคณิตศาสตร์ วิทยาศาสตร์และภาษาอังกฤษระหว่างปีการศึกษา 2552 ถึง 2556 จะมีค่าต่ำกว่าร้อยละ 50 และเป็นปัญหาที่ต้องแก้ไขอย่างเร่งด่วน แต่สำหรับโครงการวิจัยนี้เลือกวิเคราะห์เฉพาะแบบสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2550 ถึง 2552 เพื่อเป็นตัวอย่างสำหรับระดับชั้นอื่นและวิชาอื่นต่อไป

3) เนื้อหาของแบบสอบ ONET วิชาวิทยาศาสตร์สัมพันธ์กับโครงสร้างมาตรฐานการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 กล่าวคือ ข้อสอบ ONET ทุกข้อสามารถระบุได้ว่าวัดมาตรฐานการเรียนรู้และสาระการเรียนรู้ใด ทำให้คำตอบข้อสอบสัมพันธ์กับความสามารถนักเรียนตามที่ระบุในหลักสูตรด้วย ดังนั้นโครงการวิจัยนี้จึงเลือกใช้โครงสร้างมาตรฐานการเรียนรู้ วิทยาศาสตร์ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (กระทรวงศึกษาธิการ, 2551; สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2552) เป็นกรอบการวัดเพื่อประมาณค่าพารามิเตอร์ความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 และค่าพารามิเตอร์ข้อสอบ ONET ด้วยทฤษฎีการวัดของราส์ (Rasch measurement theory: Rasch, 1960; Wright & Stone, 1979) ก่อนที่จะนำค่าพารามิเตอร์ที่ได้ไปใช้กำหนดคะแนนจุดตัดขั้นต้นต่อไป

4) โครงการวิจัยนี้ใช้ข้อมูลทุติยภูมิคำตอบข้อสอบ ONET วิชาวิทยาศาสตร์ ปีการศึกษา 2550 ถึง 2552 ข้อมูลที่ใช้วิเคราะห์จึงสุ่มจากประชากรนักเรียนที่เข้าสอบในปีการศึกษา 2550 จำนวน 961,613 คน ปีการศึกษา 2551 จำนวน 953,115 คน และปีการศึกษา 2552 จำนวน 898,372 คน

5) ผู้วิจัยใช้ข้อมูลคำตอบข้อสอบและข้อสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2550 จำนวน 40 ข้อ ปีการศึกษา 2551 จำนวน 40 ข้อ และปีการศึกษา 2552 จำนวน 37 ข้อ รวมทั้งสิ้น 117 ข้อ มาวิเคราะห์คะแนนจุดตัดขั้นต้นของแต่ละปีการศึกษา โดยข้อสอบ ONET จะถูกวิเคราะห์เชิงคุณภาพเกี่ยวกับมาตรฐานการเรียนรู้หรือสาระการเรียนรู้และตัวชี้วัดที่มุ่งวัด ระดับความซับซ้อนของข้อสอบ และความสามารถของนักเรียนที่ใช้ตอบข้อสอบ แล้วนำผลวิเคราะห์เชิงคุณภาพนี้ไปใช้ร่วมกับผลวิเคราะห์เชิงปริมาณที่ได้จากคำตอบข้อสอบ ONET นอกจากนี้โครงการวิจัยต้องการตรวจสอบว่าคะแนนจุดตัดขั้นต้นจะแปรเปลี่ยนข้ามปีหรือไม่ จึงกำหนดขอบเขตของการวิจัยให้ครอบคลุมข้อสอบ ONET จำนวน 3 ปีการศึกษา ทั้งนี้ในกระบวนการเปรียบเทียบคะแนนจุดตัดขั้นต้นมีข้อตกลงเบื้องต้น 3 ประการคือ ประการแรก การทดสอบในแต่ละปีจะต้องมีบริบทและสถานการณ์ที่ไม่แตกต่างกัน ประการที่สอง แบบสอบจะต้องมีข้อกำหนดลักษณะแบบสอบ (test specification) เหมือนกัน และประการที่สาม แบบสอบในแต่ละปีถูกสร้างด้วยวัตถุประสงค์การประเมินเดียวกัน (Zieky, Perie, & Livingston, 2008) ดังนั้นถ้าคุณสมบัติแบบสอบ ONET ทั้งสามปีไม่ตรงตามข้อตกลงเบื้องต้นทั้งหมด ผลการเปรียบเทียบคะแนนจุดตัดขั้นต้นข้ามปีการศึกษาอาจจะถูกต้องแม่นยำลดลง

6) โครงการวิจัยนี้กำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ ซึ่งในปัจจุบันมีหน่วยงานที่นำวิธีนี้ไปใช้ เช่น กรมการศึกษา (Department of Education) ของรัฐแคลิฟอร์เนีย มิชิแกน โอเรกอน และเท็กซัส ในประเทศสหรัฐอเมริกา (Draney & Wilson, 2011; O'Malley, Keng, & Miles, 2012; Wyse, 2013) และกระทรวงศึกษาธิการและการวิจัย (Federal Ministry of Education and Research) ของประเทศสหพันธ์สาธารณรัฐเยอรมัน (Harsch, Pant, & Koller, 2010)

วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์มีลักษณะเด่นสองประการที่เหมาะสมกับแบบสอบ ONET คือ ประการแรก วิธีดังกล่าวเชื่อมโยงกระบวนการทดสอบเข้ากับกระบวนการกำหนดคะแนนจุดตัด โดยมีทฤษฎีจิตวิทยาการคิด (cognitive psychology) และทฤษฎีการตอบสนองข้อสอบ (item response theory) เป็นฐานรวมเรียกว่า กรอบการสร้างโมเดลภาวะสันนิษฐาน (Construct Modeling Framework: Wilson, 2005) นั้นหมายความว่าในกรณีที่การทดสอบ ONET ซึ่งไม่มี คำบรรยายระดับความสามารถของนักเรียนปรากฏอย่างเด่นชัดในขั้นตอนแรกก่อนสร้างข้อสอบ ก็สามารถใช้กรอบการสร้างโมเดลภาวะสันนิษฐานสร้างคำบรรยายระดับความสามารถขึ้นมาได้ในรูปของแผนที่ภาวะสันนิษฐาน (construct map) แล้วใช้แผนที่ดังกล่าวเป็นตัวเชื่อมข้อมูลของแบบสอบ ONET เข้ากับกระบวนการกำหนดคะแนนจุดตัด

ประการที่สอง วิธีดังกล่าวใช้ได้ดีกับแบบสอบ ONET ที่วัดความสามารถของนักเรียนหลายมาตรฐานการเรียนรู้หรือสาระการเรียนรู้ เนื่องจากวิธีดังกล่าวเชื่อมโยงเข้ากับทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (unidimensional) และพหุมิติ (multidimensional) จึงสามารถนำสารสนเทศที่ได้จากโมเดลการตอบสนองข้อสอบมาใช้ประกอบการกำหนดคะแนนจุดตัดขั้นต้นได้ ความเป็นพหุมิติ

ของแบบสอบ ONET เป็นได้ทั้งระดับข้อสอบคือข้อสอบหนึ่งข้อวัดมากกว่าหนึ่งมาตรฐานการเรียนรู้ (within-item) และระดับแบบสอบคือ ข้อสอบหนึ่งข้อวัดหนึ่งมาตรฐานการเรียนรู้ (between-item) แต่อยู่รวมกับข้อสอบอื่นๆ ซึ่งวัดมาตรฐานการเรียนรู้แตกต่างกัน ในแบบสอบฉบับเดียวกัน

7) โครงการวิจัยนี้ไม่เปรียบเทียบประสิทธิภาพของวิธีกำหนดคะแนนจุดตัดต่างๆ เนื่องจากการกำหนดคะแนนจุดตัดแต่ละวิธีมีหลักการพื้นฐานและเหมาะสมกับบริบทการประเมินที่แตกต่างกัน ทำให้คะแนนจุดตัดที่เกิดขึ้นจากคะแนนสอบชุดเดียวกันแต่ต่างวิธีกันจะมีผลลัพธ์และคำอธิบายแตกต่างกันด้วย ดังนั้นการเปรียบเทียบประสิทธิภาพของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์กับวิธีอื่นๆ จึงน่าสนใจน้อยกว่าการประยุกต์ใช้วิธีดังกล่าวกับการทดสอบระดับชาติของประเทศไทย

โดยทั่วไปแล้ววิธีกำหนดคะแนนจุดตัดส่วนใหญ่จะใช้แนวคิดแบบดั้งเดิมคือ จะกำหนดคะแนนจุดตัดภายหลังจากที่ได้สร้างสเกลคะแนนสอบขึ้นแล้ว จากนั้นผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจจะร่วมกันอภิปรายความเหมาะสมของค่าที่จะใช้เป็นคะแนนจุดตัด อย่างไรก็ตามคะแนนจุดตัดที่เกิดขึ้นมักไม่คงเส้นคงวา (Skaggs, 2013) และไม่สอดคล้องกับคำบรรยายระดับความสามารถของนักเรียนที่ทำแบบสอบฉบับนั้นๆ เนื่องจากคำบรรยายระดับความสามารถไม่ได้ถูกใช้ในกระบวนการสร้างแบบสอบ (Perie, 2008) และสารสนเทศสำคัญของแบบสอบไม่ได้ถูกนำมาพิจารณาด้วย ส่วนแนวคิดใหม่ที่พยายามเชื่อมโยงกระบวนการกำหนดคะแนนจุดตัดเข้ากับกระบวนการทดสอบมีเป้าหมายหลักเพื่อให้คะแนนจุดตัดที่กำหนดขึ้นเป็นที่ยอมรับ ใช้งานได้และนำไปใช้ได้ตรงตามวัตถุประสงค์ ตัวอย่างวิธีกำหนดคะแนนจุดตัดแบบแนวคิดใหม่นี้เช่นวิธีของ Mislevy, Steinberg และ Almond (2003) วิธีของ Wang (2003) วิธีของ Reckase (1998) และวิธีของ Wilson และ Draney (2002) ในทำนองเดียวกันโครงการวิจัยนี้ก็ไม่ได้สื่อความว่าวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ (Wilson & Draney, 2002) ดีกว่าวิธีอื่นๆ ในกลุ่มเดียวกัน แต่ต้องการนำเสนอวิธีการที่เหมาะสมกับบริบทการทดสอบของประเทศไทยอีกวิธีหนึ่ง

นิยามศัพท์เฉพาะ

โครงการวิจัยนี้มีศัพท์เฉพาะที่ไม่คุ้นเคยหลายคำ ศัพท์บางคำไม่มีชื่อเรียกเป็นภาษาไทย ในขณะที่บางคำแม้ว่าจะเป็นศัพท์ที่ใช้กันโดยทั่วไปแต่ก็มีความหมายเฉพาะในรายงานฉบับนี้ ดังนั้นผู้วิจัยจึงนิยามศัพท์เฉพาะมีดังต่อไปนี้

การกำหนดคะแนนจุดตัดความสามารถทางวิทยาศาสตร์ เป็นกระบวนการระบุค่าเพื่อจำแนกนักเรียนออกเป็นกลุ่มตามระดับความสามารถจากคะแนนสอบ ONET โดยคะแนนจุดตัดจะสัมพันธ์กับคำบรรยายระดับความสามารถทางวิทยาศาสตร์ (performance level descriptor) ซึ่งถูกกำหนดขึ้นก่อนกระบวนการกำหนดคะแนนจุดตัด คำว่า การกำหนดคะแนนจุดตัด ตรงกับคำภาษาอังกฤษว่า cutscore setting หรือ cut-score setting และสามารถใช้แทนกันได้กับคำว่า การกำหนดมาตรฐาน (standard setting) และการกำหนดมาตรฐานความสามารถ (performance standard setting)

(Cizek, 2012; Zieky, 2012) แต่ในขณะเดียวกัน การกำหนดคะแนนจุดตัดก็สามารถพิจารณาเป็นขั้นตอนย่อยของกระบวนการกำหนดมาตรฐานความสามารถ (Cizek & Bunch, 2007) สำหรับโครงการวิจัยนี้เลือกใช้คำว่า คะแนนจุดตัด เพื่อหลีกเลี่ยงความสับสนและซ้ำซ้อนกับความหมายอื่นๆ ของคำว่ามาตรฐาน เช่น มาตรฐานการเรียนรู้ (learning standard) มาตรฐานความสามารถ (performance standard) มาตรฐานเนื้อหา (content standard) และมาตรฐานการทดสอบ (testing standard) ที่ปรากฏในรายงานฉบับนี้ และมีมุมมองว่าการกำหนดคะแนนจุดตัดเป็นกระบวนการย่อยของการกำหนดมาตรฐานความสามารถ

คะแนนจุดตัดขั้นต้น หมายถึงค่าที่ได้จากการวิเคราะห์ข้อมูลคำตอบข้อสอบ ONET ตามทฤษฎีการวัดของราสส์และผู้วิจัยเสนอให้ค่าเหล่านั้นสามารถใช้เป็นคะแนนจุดตัดได้โดยไม่ขัดแย้งกับคำบรรยายระดับความสามารถทางวิทยาศาสตร์ที่กำหนดขึ้นและสามารถจำแนกนักเรียนเป็นกลุ่มได้ ดังนั้นคะแนนจุดตัดขั้นต้นอาจไม่ใช่คะแนนจุดตัดจริง โดยผู้มีอำนาจจะเป็นผู้ตัดสินค่าที่เหมาะสมที่สุด

แผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ เป็นโครงสร้างทางทฤษฎีที่ผู้วิจัยสันนิษฐานขึ้นในเบื้องต้นจากผลวิเคราะห์คุณสมบัติข้อสอบ ONET วิชาวิทยาศาสตร์โดยผู้เชี่ยวชาญด้านเนื้อหาวิทยาศาสตร์ ซึ่งอาจเรียกว่า แผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์เชิงทฤษฎี (theoretical construct map) จากนั้นโครงสร้างดังกล่าวจะถูกตรวจสอบความถูกต้องในภายหลังด้วยทฤษฎีการวัดของราสส์ทำให้ได้แผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์เชิงประจักษ์ (empirical construct map) คำว่า แผนที่ภาวะสันนิษฐาน ตรงกับคำภาษาอังกฤษว่า construct map ซึ่งจะทำหน้าที่เหมือนคำบรรยายระดับความสามารถที่ใช้กันทั่วไปในกระบวนการกำหนดคะแนนจุดตัด แต่มีคุณสมบัติที่เหนือกว่าคือ นอกจากจะแบ่งความสามารถทางวิทยาศาสตร์ของนักเรียนเป็นระดับที่ชัดเจนมากขึ้นบนพื้นฐานของข้อมูลเชิงประจักษ์และแสดงโครงสร้างความสามารถแบบลดหลั่นกันจากระดับมากที่สุดระดับน้อยอย่างเป็นระบบแล้ว ยังเชื่อมโยงทุกขั้นตอนในกระบวนการทดสอบเข้ากับทุกขั้นตอนในกระบวนการกำหนดคะแนนจุดตัดด้วย

โมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์แบบสุ่มจากการวัดแบบหลายมิติ ตรงกับคำภาษาอังกฤษว่า Multidimensional Random Coefficients Multinomial Logit model หรือโมเดล MRCML เป็นโมเดลทั่วไปทางคณิตศาสตร์ที่ใช้อธิบายมโนทัศน์การออกแบบการวัดทุกประเภทตามทฤษฎีการวัดของราสส์ ดังนั้นไม่ว่าการออกแบบการวัดจะมีลักษณะอย่างไร เช่น คุณลักษณะแฝงเป็นแบบเอกมิติ (unidimensional) หรือพหุมิติ (multidimensional) การให้คะแนนคำตอบจะเป็นแบบสองค่า (dichotomous) หรือหลายค่า (polytomous) รูปแบบข้อสอบจะเป็นข้อสอบเดี่ยวหรือเป็นข้อสอบกลุ่ม (item bundle) หรือจะนำคุณสมบัติของผู้สอบ (person properties) หรือคุณสมบัติข้อสอบ (item properties) มาวิเคราะห์ร่วมด้วยหรือไม่ก็ตาม ก็สามารถประมาณค่าพารามิเตอร์ต่างๆ ตามทฤษฎีการวัดของราสส์ได้ทั้งหมด

วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ ตรงกับคำภาษาอังกฤษว่า construct mapping method เป็นแนวคิดและวิธีการที่เชื่อมโยงกระบวนการกำหนดคะแนนจุดตัดเข้ากับ

กระบวนการทดสอบโดยมีแผนที่ภาวะสันนิษฐานเป็นแกนกลางเชื่อมโยงสารสนเทศของข้อสอบ ONET วิชาวิทยาศาสตร์ ความสามารถแฝงทางวิทยาศาสตร์ของนักเรียนที่ใช้ตอบข้อสอบ คำตอบที่นักเรียนเลือกตอบและการให้คะแนนคำตอบเหล่านั้น มาใช้เป็นหลักฐานประกอบการตัดสินใจตัดสินคะแนนจุดตัด ดังนั้นก่อนเริ่มกระบวนการกำหนดคะแนนจุดตัด แผนที่ภาวะสันนิษฐานจะถูกตรวจสอบความถูกต้องด้วยข้อมูลคำตอบข้อสอบ ONET จึงทำให้แผนที่ภาวะสันนิษฐานจากเดิมซึ่งเป็นโครงสร้างเชิงทฤษฎีกลายเป็นโครงสร้างเชิงประจักษ์ แล้วจึงนำไปใช้ในกระบวนการกำหนดคะแนนจุดตัดต่อไป ผู้วิจัยจึงเรียกวิธีการกำหนดคะแนนจุดตัดนี้ว่า วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์

ประโยชน์ที่ได้รับ

โครงการวิจัยนี้เป็นประโยชน์ต่อกระบวนการทดสอบและกระบวนการกำหนดคะแนนจุดตัด ผลวิจัยที่เกิดขึ้นจะถูกต้องและน่าเชื่อถือมากขึ้นตามนิยามของคณะกรรมการวิจัยแห่งชาติประเทศสหรัฐอเมริกา (NRC, 2001) เนื่องจากผลวิจัยจะสัมพันธ์กับโครงสร้างภายในของสิ่งที่สนใจหรือโมเดลตัวแปรแฝง และสอดคล้องอย่างยิ่งกับมาตรฐานการทดสอบทางการศึกษาและจิตวิทยาของประเทศสหรัฐอเมริกา (AERA, APA, & NCME, 2014) ประโยชน์ที่ได้รับมีรายละเอียดดังนี้

- 1) ได้แนวทางกำหนดคะแนนจุดตัดที่เชื่อมโยงกับกระบวนการสร้างข้อสอบและได้แนวทางส่งเสริมความเชี่ยวชาญด้านการออกแบบข้อสอบ (item design) ให้กับผู้ออกข้อสอบ ONET ของสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) ประโยชน์ข้อนี้เป็นประโยชน์ตั้งต้นของการนำโมเดลทักษะการคิด (cognition model) หรือโมเดลตัวแปรแฝง (latent variable model) มาใช้ในกระบวนการพัฒนาการทดสอบแล้วขยายไปสู่การกำหนดคะแนนจุดตัด ซึ่งแผนที่ภาวะสันนิษฐานนี้ถือว่าเป็นรูปแบบหนึ่งของโมเดลตัวแปรแฝงที่นอกจากจะระบุส่วนประกอบของสิ่งที่สนใจแล้วยังบรรยายลักษณะภายในของสิ่งที่สนใจนั้นด้วย ผู้ออกข้อสอบ ONET จะมีแนวทางที่ชัดเจนมากขึ้นว่าสิ่งที่ต้องการจะวัดนั้นคืออะไรเพื่อสร้างข้อสอบที่ให้ข้อมูลที่เป็นประโยชน์ต่อการกำหนดคะแนนจุดตัด แทนที่จะสนใจเฉพาะจำนวนข้อสอบที่นักเรียนตอบถูกหรือคะแนนรวมจากแบบสอบทั้งฉบับ โดยนอกเหนือจากที่ผู้ออกข้อสอบ ONET จะระบุว่าข้อสอบแต่ละข้อนั้นวัดมาตรฐานการเรียนรู้และสาระการเรียนรู้อะไรแล้ว ตัวเลือกหรือแม้กระทั่งคำตอบที่นักเรียนตอบเองจะสามารถตีความได้ว่านักเรียนมีความสามารถทางวิทยาศาสตร์อยู่ในระดับใดตามแผนที่ภาวะสันนิษฐานที่พัฒนาขึ้น ซึ่งเท่ากับว่าเมื่อนำแผนที่ภาวะสันนิษฐานนี้มาใช้ในกระบวนการกำหนดคะแนนจุดตัด ค่าของคะแนนจุดตัดขั้นต้นจะเชื่อมโยงเข้ากับมาตรฐานการเรียนรู้และสาระการเรียนรู้ที่ข้อสอบ ONET ต้องการวัดและความสามารถที่แตกต่างกันของนักเรียน คำตอบหรือตัวเลือกที่นักเรียนเลือกจะสัมพันธ์กับระดับความสามารถของนักเรียนซึ่งเป็นหลักฐานยืนยันความสามารถของนักเรียนเอง ตรรกะการสร้างข้อสอบที่เน้นโมเดลตัวแปรแฝงไม่ได้มุ่งสร้างตัวเลือกเพื่อลวงนักเรียน แต่เป็นการสร้างตัวเลือกที่เหมาะสมกับความสามารถที่แตกต่างกันของนักเรียน เมื่อผู้ออกข้อสอบใช้แผนที่ภาวะสันนิษฐานสร้าง

ข้อสอบ ONET วิทยาศาสตร์ก็จะเกิดความเชี่ยวชาญมากขึ้น สามารถระบุคุณลักษณะหรือความสามารถที่เป็นไปได้ของนักเรียนได้แม่นยำมากขึ้นและเมื่อนำสารสนเทศเหล่านี้ไปใช้กำหนดคะแนนก็จะทำให้คะแนนจุดตัดถูกต้องและน่าเชื่อถือมากขึ้นเช่นกัน

ในกระบวนการสร้างแบบสอบ ONET ผู้ออกข้อสอบจะต้องระบุว่าข้อสอบแต่ละข้อวัดมาตรฐานการเรียนรู้และสาระการเรียนรู้อะไร ความถูกต้องแม่นยำของข้อกำหนดลักษณะแบบสอบ (test specification) จึงสำคัญอย่างมากต่อการวิเคราะห์คะแนนสอบ รวมทั้งสะท้อนความเชี่ยวชาญของผู้ออกข้อสอบทั้งในด้านเนื้อหาวิทยาศาสตร์และความเข้าใจการประเมินการเรียนรู้ ผลการตรวจสอบความเป็นพหุมิติของแบบสอบ ONET สามารถใช้เป็นหลักฐานยืนยันความถูกต้องของข้อกำหนดลักษณะแบบสอบซึ่งส่งผลต่อการกำหนดคะแนนจุดตัดความสามารถรายด้าน ตัวอย่างเช่น สมมติว่าผู้ออกข้อสอบมีความเห็นว่าข้อสอบ ONET วิชาวิทยาศาสตร์ทุกข้อวัดมาตรฐานการเรียนรู้ 8.1 เกี่ยวกับทักษะกระบวนการทางวิทยาศาสตร์ แต่ผลวิเคราะห์กลับแสดงว่าข้อสอบทุกข้อไม่ได้เป็นเช่นนั้น เนื่องจากโมเดลพหุมิติไม่กลมกลืนกับคำตอบข้อสอบของนักเรียน สารสนเทศในลักษณะนี้จะช่วยให้ผู้ออกข้อสอบกลับมาทบทวนกระบวนการออกแบบข้อสอบของตนเองให้สามารถวัดได้ตรงตามที่ต้องการมากขึ้น และไม่ใช้คะแนนจุดตัดตัดสินทักษะกระบวนการทางวิทยาศาสตร์ของนักเรียน

2) สร้างคะแนนจุดตัดจากข้อสอบจำนวนน้อยแต่ให้สารสนเทศเกี่ยวกับความสามารถของนักเรียนมาก เนื่องจากมาตรฐานการเรียนรู้วิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 มีจำนวนถึง 12 มาตรฐาน การสร้างข้อสอบ ONET เพื่อวัดทุกมาตรฐานการเรียนรู้จึงต้องใช้ข้อสอบจำนวนหลายข้อและข้อสอบเหล่านั้นจะต้องให้ข้อมูลเพื่อยืนยันความถูกต้องของคะแนนจุดตัดด้วย แต่อย่างไรก็ตาม ภายใต้กรอบโมเดล MRCML ผู้ออกข้อสอบ ONET วิชาวิทยาศาสตร์จะสามารถออกแบบข้อสอบให้วัดมากกว่าหนึ่งมาตรฐานการเรียนรู้ (หรือมากกว่า 1 สาระการเรียนรู้) คำตอบข้อสอบจึงถูกนำไปใช้ประมาณความสามารถของนักเรียนได้อย่างคุ้มค่า ทำให้จำนวนข้อสอบที่วัดเพียงหนึ่งมาตรฐานการเรียนรู้ลดลง และในขณะเดียวกัน ข้อมูลความสามารถของนักเรียนก็สามารถรวบรวมได้จากข้อสอบอื่นที่ไม่จำเป็นต้องสร้างมาเพื่อวัดมาตรฐานการเรียนรู้นั้นเพียงอย่างเดียว ซึ่งเท่ากับว่าผู้ออกข้อสอบ ONET จะมีข้อมูลจากแหล่งต่างๆ มาใช้ประมาณความสามารถทางวิทยาศาสตร์ของนักเรียนได้แม่นยำมากขึ้น ดังนั้นเมื่อกำหนดคะแนนจุดตัดจากค่าประมาณความสามารถดังกล่าวก็จะสามารถกำหนดคะแนนจุดตัดรายด้านได้จากข้อสอบจำนวนน้อยได้ด้วย

3) ติความคะแนนจุดตัดและคะแนนสอบด้วยคะแนนสเกล (scale score) ที่ได้จากทฤษฎีการตอบสนองข้อสอบ เนื่องจากผลการวิเคราะห์ข้อสอบและคะแนนสอบ ONET ที่ผ่านมาใช้หลักการของทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) ซึ่งมีข้อจำกัดที่สำคัญคือ ความแปรเปลี่ยนของคะแนนสอบตามกลุ่มผู้สอบและข้อสอบ ทำให้เมื่อนำคะแนนสอบไปกำหนดเป็นคะแนนจุดตัด ก็จะทำให้คะแนนจุดตัดขึ้นต้นขึ้นอยู่กับกลุ่มนักเรียนที่เข้าสอบ และจำนวนข้อสอบ ONET ที่ตอบถูกต้อง ส่วนหนึ่งของผลการวิจัยในครั้งนี้ เป็นการนำเสนอคะแนนสเกลที่เกิดขึ้นจากการนำค่าประมาณความสามารถแฝงของนักเรียนจากทฤษฎีการตอบสนองข้อสอบโดยใช้โมเดลการวัด

ของราล์ฟมาแปลงเป็นคะแนนสเกลที่มีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 โดยที่คะแนนสเกลนี้ยังมีคุณสมบัติความไม่แปรเปลี่ยนเช่นเดิม ประโยชน์ของคะแนนสเกลนอกจากใช้แสดงความสามารถของนักเรียนที่ถูกต้องมากยิ่งขึ้นเมื่อเทียบกับคะแนนรวมดิบแล้ว ยังช่วยให้ผู้ใช้คะแนนสอบ ONET เข้าใจค่าประมาณความสามารถแฝงได้ง่ายขึ้นเพราะคะแนนจุดตัดขั้นต้นและค่าประมาณความสามารถแฝงเดิมมีทั้งค่าบวกและค่าลบ อาจทำให้ผู้ใช้คะแนนสอบ ONET ตีความหมายของคะแนนคลาดเคลื่อนได้

4) ได้คะแนนจุดตัดขั้นต้นที่ให้สารสนเทศเชิงวินิจฉัยความสามารถทางวิทยาศาสตร์ของนักเรียน เนื่องจากโมเดลเอกมิติสมมติให้ข้อสอบทุกข้อในแบบสอบวัดสิ่งที่สนใจเพียงมิติเดียว ทำให้ค่าประมาณความสามารถแฝงที่ได้จากโมเดลเอกมิติเป็นภาพรวม และยังฝ่าฝืนธรรมชาติของความสามารถทางวิทยาศาสตร์ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ที่ประกอบด้วยมาตรฐานการเรียนรู้และสาระการเรียนรู้หลายด้าน ผลที่จะเกิดขึ้นจากโมเดลพหุมิติของโครงการวิจัยนี้จะให้ค่าประมาณความสามารถแฝงรายด้านจึงสามารถกำหนดคะแนนจุดตัดขั้นต้นรายด้านได้ด้วย สารสนเทศที่จะได้รับจึงชัดเจนมากขึ้น มีแนวทางการตีความคะแนนของนักเรียนที่สูงกว่าคะแนนจุดตัดขั้นต้นว่ายังมีจุดด้อยอะไรบ้าง นำไปสู่การวินิจฉัยความสามารถทางวิทยาศาสตร์ของนักเรียนทั้งในระดับประเทศ เขตพื้นที่การศึกษาและโรงเรียน

5) ให้ข้อมูลที่ถูกต้องประกอบการกำหนดมาตรฐานความสามารถทางวิทยาศาสตร์ โดยในเบื้องต้น ผู้วิจัยใช้แผนที่ภาวะสันนิษฐานเพื่อกำหนดจุดตัดที่จำแนกนักเรียนตามระดับความสามารถเป็นสองระดับ ข้อดีของวิธีดังกล่าวคือมีหลักการเช่นเดียวกับทฤษฎีการวัดของราล์ฟและใช้ประโยชน์จากผลวิเคราะห์จากทฤษฎีดังกล่าวกำหนดคะแนนจุดตัดขั้นต้นได้อย่างมีประสิทธิภาพ ทำให้คะแนนจุดตัดที่กำหนดขึ้นไม่แปรเปลี่ยนตามแบบสอบและกลุ่มผู้สอบ ผลวิจัยที่เกิดขึ้นจึงเป็นข้อมูลอย่างดีสำหรับผู้บริหารการศึกษานำไปใช้ประกอบการตัดสินใจเมื่อต้องนำคะแนนจุดตัดไปใช้สร้างมาตรฐานความสามารถทางวิทยาศาสตร์ และประเมินผลสัมฤทธิ์การเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551

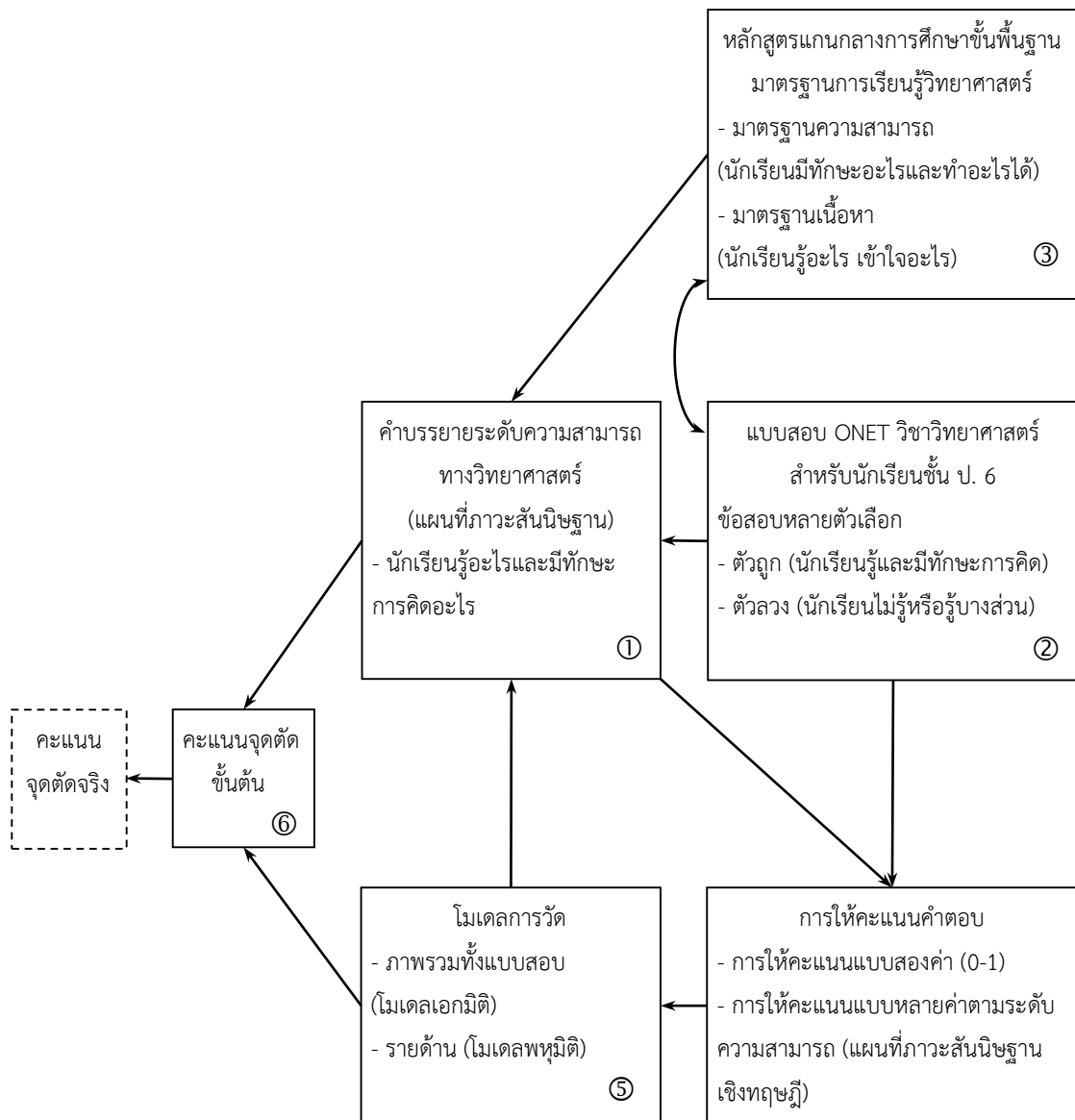
กรอบแนวคิดของการวิจัย

ผู้วิจัยเลือกวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์เพื่อกำหนดคะแนนจุดตัดขั้นต้นจากคะแนนสอบ ONET เนื่องจากในกระบวนการสร้างแบบสอบ ONET นั้น คำบรรยายระดับความสามารถทางวิทยาศาสตร์ซึ่งถือว่าเป็นหัวใจสำคัญของการกำหนดคะแนนจุดตัดและเป็นองค์ประกอบสำคัญของการพัฒนาระบบการประเมินแนวใหม่ตามนิยามของคณะกรรมการวิจัยแห่งชาติของประเทศสหรัฐอเมริกาไม่ปรากฏชัดเจน ทำให้คะแนนสอบ ONET อาจจะไม่สามารถตีความได้ตามโครงสร้างของสิ่งที่ต้องการวัดและอาจจะนำไปใช้ไม่ตรงตามวัตถุประสงค์ ดังนั้นจึงจำเป็นต้องพัฒนาคำบรรยายระดับความสามารถทางวิทยาศาสตร์ขึ้น โดยอาศัยกรอบการสร้างโมเดล

ภาวะสันนิษฐาน คำบรรยายระดับความสามารถที่เกิดขึ้นจะอยู่ในรูปของแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ ซึ่งเกิดจากการวิเคราะห์ข้อมูลข้อสอบ ONET แต่ละข้อทั้งคำถามและตัวเลือก (ดู ② ของภาพที่ 1.1) และคำบรรยายมาตรฐานการเรียนรู้วิทยาศาสตร์ในหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2551 และตัวชี้วัดต่างๆ (ดู ③ ของภาพที่ 1.1)

สารสนเทศของข้อสอบ ONET และมาตรฐานการเรียนรู้จะถูกนำมาใช้สร้างคำบรรยายระดับความสามารถทางวิทยาศาสตร์หรือแผนที่ภาวะสันนิษฐาน (ดู ① ของภาพที่ 1.1) สาเหตุที่หมายเลขแสดงลำดับความคิดของนักวิจัยที่เรียงจาก ② ③ และ ① แทนที่จะเป็น ① ② และ ③ เนื่องจากกระบวนการพัฒนาแบบสอบ ONET ที่เป็นอยู่จริงนั้นไม่มีคำบรรยายความสามารถมาก่อน ทำให้โครงสร้างความสามารถทางวิทยาศาสตร์ซึ่งเป็นเป้าหมายที่แท้จริงของการพัฒนาแบบสอบ ONET ไม่ชัดเจน จึงจำเป็นต้องใช้แนวคิดแบบย้อนกลับเพื่อสร้างคำบรรยายระดับความสามารถก่อน แนวคิดแบบย้อนกลับนี้จึงไม่เหมือนกับกระบวนการพัฒนาแบบสอบที่ควรจะเป็น คำบรรยายดังกล่าวไม่ควรพัฒนาขึ้นโดยใช้มาตรฐานการเรียนรู้เพียงอย่างเดียวเพราะจะทำให้คำบรรยายนั้นไม่สอดคล้องกับข้อสอบ ONET ที่สร้างเสร็จเรียบร้อยแล้ว ในทำนองเดียวกันการสร้างคำบรรยายโดยพิจารณาเฉพาะแบบสอบ ONET ก็อาจจะไม่สมเหตุสมผล เนื่องจากแบบสอบ ONET มีมาตรฐานการเรียนรู้ตามที่ปรากฏในหลักสูตรเป็นฐานความคิดสำคัญ (ดูขอบเขตการวิจัยข้อ 3 ประกอบ) แต่อย่างไรก็ตามแบบสอบ ONET ก็ไม่ได้มีรายละเอียดเหมือนกันทุกประการกับมาตรฐานการเรียนรู้ จึงทำให้ส่วนประกอบทั้งสองส่วนแยกออกจากกัน (แต่สัมพันธ์กัน) และส่งผลต่อคำบรรยายระดับความสามารถที่พัฒนาขึ้น

เมื่อพัฒนาร่างคำบรรยายระดับความสามารถเรียบร้อยแล้ว สารสนเทศนี้และข้อสอบ ONET จะถูกใช้เป็นเกณฑ์การให้คะแนนคำตอบ (ดู ④ ของภาพที่ 1.1) แม้ว่าในความเป็นจริงนั้นข้อสอบ ONET จะให้คะแนนแบบสองค่า (dichotomous scoring) คือตอบถูกได้ 1 หรือคะแนนเต็มของข้อนั้น และได้ 0 คะแนนเมื่อตอบแบบอื่น แต่ก็ไม่สามารถสรุปว่าความสามารถของนักเรียนที่แบ่งออกเป็นสองกลุ่มแบบนี้จะไม่ผันแปรภายในกลุ่ม กล่าวคือนักเรียนที่ตอบถูกข้อสอบยากจะมีความสามารถมากกว่านักเรียนที่ตอบถูกข้อสอบง่าย และนักเรียนที่เลือกตัวलगต่างกันในเรื่องเดียวกันก็ไม่สามารถสรุปว่ามีความสามารถเท่ากัน ข้อสอบหลายตัวเลือกที่ให้คะแนนคำตอบแบบสองค่ามีข้อตกลงเบื้องต้นพื้นฐานคือ ตัวलगทั้งหมดมีระดับความไม่ถูกต้องไม่แตกต่างกัน (Lord, 1980) นอกจากนี้ข้อสอบดังกล่าวยังไม่สามารถรักษาลักษณะเฉพาะของตัวलगที่ไม่เหมือนกันได้อีกด้วย (van der Linden & Hambleton, 1997) ทำให้สารสนเทศข้อสอบสูญหายไปเช่นกัน ดังนั้นผู้วิจัยเชื่อว่าการให้คะแนนคำตอบ ONET กับระดับความสามารถของนักเรียนจึงไม่ใช่เรื่องเดียวกันแต่สอดคล้องกัน กล่าวคือนักเรียนอาจจะมีความสามารถแตกต่างกันแต่ได้คะแนนเท่ากันได้ แนวคิดการวิจัยแบบนี้จึงเป็นความพยายามสกัดสารสนเทศที่มีอยู่ของข้อสอบ ONET ออกมาแล้วสร้างเป็นแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์



ภาพที่ 1.1 กรอบแนวคิดของวิเคราะห์คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์

คำตอบข้อสอบจะถูกให้คะแนนใหม่ตามแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์และจะนำไปวิเคราะห์ตามโมเดลการวัดต่อไป คำตอบของนักเรียนจะถูกวิเคราะห์ด้วยโมเดลการวัด 2 ประเภท (ดู ⑤ ในภาพที่ 1.1) ตามข้อกำหนดลักษณะแบบสอบ ONET คือโมเดลการวัดของราส์แบบเอกมิติและแบบพหุมิติ สาเหตุที่ต้องเลือกใช้โมเดลการวัดของราส์ เนื่องจากหลักการที่ใช้สร้างแผนที่ภาวะสันนิษฐานเป็นการกำหนดตำแหน่งความสามารถของนักเรียนอย่างคร่าวๆ เป็นหลักการเดียวกับพารามิเตอร์ความสามารถแฝงของผู้สอบและพารามิเตอร์ความยากที่ได้จากโมเดลการวัดทั้งสองประเภท

พารามิเตอร์ความยากข้อสอบ ไม่ว่าจะเป็นจากการให้คะแนนแบบสองค่าหรือแบบหลายค่า จะถูกเปรียบเทียบกันในรูปแบบของแผนภาพการแจกแจงพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบ จากนั้นจึงปรับปรุงแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ที่ได้ออกแบบไว้ในตอนต้นบนพื้นฐานของข้อมูลเชิงประจักษ์แล้วดำเนินการวิเคราะห์ซ้ำ (ดูลูกศรที่ชี้จาก ① ไป ④) จนได้แผนที่ภาวะสันนิษฐานเชิงประจักษ์ที่จำแนกนักเรียนตามระดับความสามารถได้ชัดเจนมากขึ้น แล้วจึงนำผลวิเคราะห์ไปกำหนดคะแนนจุดตัดขั้นต้น (ดู ⑥ ในภาพที่ 1.1) โดยพิจารณาจากค่าพารามิเตอร์ความสามารถ ค่าพารามิเตอร์ความยากข้อสอบ ค่าความคลาดเคลื่อนของการประมาณ และการแจกแจงของค่าพารามิเตอร์ที่เกี่ยวข้อง ทั้งนี้หลังจากที่กำหนดคะแนนจุดตัดขั้นต้นแล้ว ผู้วิจัยจะตรวจสอบผลกระทบที่เกิดขึ้นจากการนำคะแนนจุดตัดไปใช้ โดยพิจารณาจากร้อยละของจำนวนนักเรียนที่มีคะแนนอย่างน้อยเท่ากับและต่ำกว่าคะแนนจุดตัดขั้นต้น ซึ่งผลวิเคราะห์ส่วนนี้จะเป็นข้อมูลเพิ่มเติมให้กับผู้มีอำนาจตัดสินใจใช้ประกอบการตัดสินใจว่าค่าใดควรจะเป็นคะแนนจุดตัดจริง ซึ่งอยู่นอกเหนือขอบเขตของโครงการวิจัยนี้ (ดูขอบเขตการวิจัยข้อ 1 ประกอบ)

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

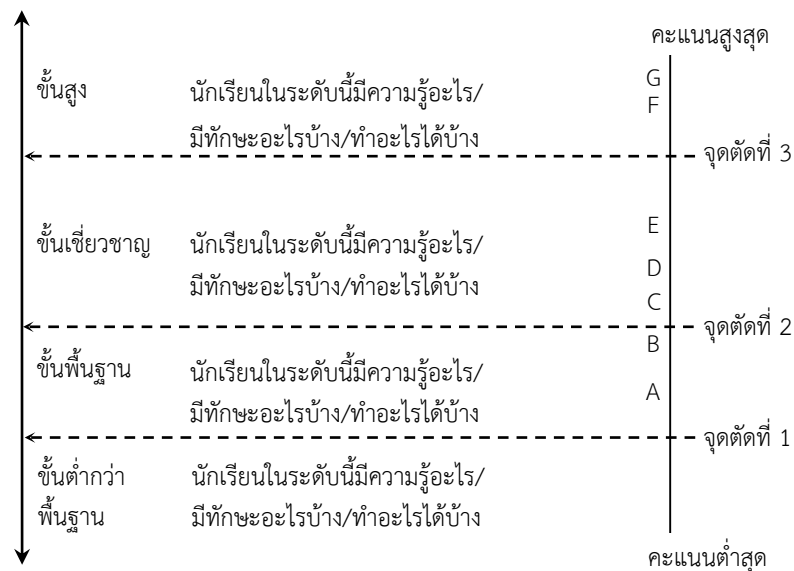
เอกสารและงานวิจัยที่เกี่ยวข้องกับโครงการวิจัยนี้มีเนื้อหาครอบคลุมแนวคิดและหลักการสำคัญของการกำหนดคะแนนจุดตัดและหลักการวิเคราะห์ข้อมูลเพื่อให้ได้คะแนนจุดตัดขั้นต้น โดยมีเนื้อหาแบ่งส่วนย่อย ได้แก่ (1) หลักการของการกำหนดคะแนนจุดตัด (2) คุณลักษณะของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ (3) แผนที่ภาวะสันนิษฐานและคำบรรยายระดับความสามารถ (4) จุดเด่นของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ (5) แผนที่ภาวะสันนิษฐานและโมเดลการวัดของราสส์ (6) การสร้างคำบรรยายระดับความสามารถด้วยวิธีทำแผนที่ภาวะสันนิษฐาน (7) คุณลักษณะของโมเดลเอกมิตีและโมเดลพหุมิติความสามารถนักเรียน และ (8) ตัวอย่างงานวิจัยที่เกี่ยวข้อง รายละเอียดของเนื้อหาจะย่อมีดังนี้

หลักการของการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัดตรงกับคำภาษาอังกฤษว่า cutscore setting หรือ standard setting หรือ performance standard setting ซึ่งคำเหล่านี้ใช้คู่กับคำว่า on test นั้นหมายความว่า หลักการกำหนดคะแนนจุดตัดไม่ได้มีเพียงการเลือกวิธีที่เหมาะสม แต่รวมถึงแบบสอบที่ใช้เป็นเครื่องมือรวบรวมข้อมูลคำตอบข้อสอบและการให้คะแนนคำตอบ แล้วจึงนำคะแนนสอบไปใช้กำหนดคะแนนจุดตัด คะแนนสอบและคะแนนจุดตัดจึงสัมพันธ์กันในรูปของผลที่ได้จากการวัดด้วยแบบสอบ และเป็นค่าแทนปริมาณหรือระดับความสามารถของนักเรียน ดังนั้นคะแนนสอบและคะแนนจุดตัดจึงต้องตีความได้ตามวัตถุประสงค์การสร้างแบบสอบนั้นด้วย (Cizek, 2012; Kane, 2006; Kane & Tannenbaum, 2013)

การกำหนดคะแนนจุดตัดจะเกี่ยวข้องกับมาตรฐานความสามารถ (performance standard) ซึ่งแตกต่างจากมาตรฐานเนื้อหา (content standard) (Cizek, 2012) กล่าวคือมาตรฐานเนื้อหามุ่งบรรยายผลการเรียนรู้ที่คาดหวังจากการเรียนรู้ของนักเรียนในเรื่องใดเรื่องหนึ่ง เช่น ความรู้ที่นักเรียนต้องเรียนและทักษะที่นักเรียนต้องมีการเรียนรู้ที่ชัดเจน ในขณะที่มาตรฐานความสามารถมุ่งระบุลักษณะและรายละเอียดความสามารถของนักเรียนที่ใช้ในการเรียนรู้ แล้วจึงจำแนกความสามารถออกเป็นระดับต่างๆ ด้วยคะแนนจุดตัด ในบางสถานการณ์คะแนนจุดตัดจะหมายถึงคะแนนการผ่าน (passing score) หรือระดับผลสัมฤทธิ์ขั้นต่ำ (minimum achievement level) แต่อย่างไรก็ตามเมื่อพิจารณาวัตถุประสงค์ของคะแนนจุดตัดพบว่า หน้าที่หลักของคะแนนจุดตัดคือจำแนกความสามารถของนักเรียนออกเป็นระดับต่างๆ ซึ่งอาจจะมีมากกว่าสองระดับก็ได้ ตัวอย่างเช่น ผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับการศึกษาขั้นพื้นฐาน (K-12) ของประเทศสหรัฐอเมริกา

แบ่งออกเป็นสี่ระดับคือ ขั้นต่ำกว่าพื้นฐาน (Below basic) ขั้นพื้นฐาน (Basic) ขั้นเชี่ยวชาญ (Proficient) และขั้นสูง (Advanced) นั้นหมายความว่า คะแนนจุดตัดระดับผลสัมฤทธิ์ทางการเรียน จะมีสามค่า ค่าแรกจะแบ่งระหว่างขั้นต่ำกว่าพื้นฐานกับขั้นพื้นฐาน ค่าที่สองจะแบ่งระหว่างขั้นพื้นฐานกับขั้นเชี่ยวชาญ และค่าที่สามจะแบ่งระหว่างขั้นเชี่ยวชาญกับขั้นสูง ตัวอย่างดังกล่าวนี้ชี้ให้เห็นว่า คะแนนจุดตัดมีความหมายมากกว่าคะแนนการผ่านเพราะผู้มีอำนาจตัดสินใจจะให้คะแนนการผ่าน เท่ากับค่าที่แบ่งระหว่างขั้นพื้นฐานกับขั้นเชี่ยวชาญก็ได้ และแตกต่างจากระดับผลสัมฤทธิ์ขั้นต่ำซึ่ง อาจจะมีค่าเท่ากับค่าที่แบ่งระหว่างขั้นต่ำกว่าพื้นฐานและขั้นพื้นฐานก็ได้



ภาพที่ 2.1 ความสัมพันธ์ระหว่างคะแนนจุดตัดและมาตรฐานความสามารถ

ความสัมพันธ์ระหว่างคะแนนจุดตัดกับมาตรฐานความสามารถดังภาพที่ 2.1 สามารถอธิบายได้ด้วยแนวคิดของแผนที่ภาวะสันนิษฐาน (construct map) ที่อยู่ภายใต้กรอบการสร้างโมเดลภาวะสันนิษฐาน (Construct Modeling framework: Wilson, 2005) กล่าวคือ เส้นตรงแนวตั้งที่มีหัวลูกศรทั้งสองด้านซ้ายมือเป็นเส้นสมมติเชิงทฤษฎีที่แสดงถึงความต่อเนื่องของการมีความรู้หรือทักษะหรือความสามารถในเรื่องใดเรื่องหนึ่งจากระดับน้อย (หัวลูกศรด้านล่าง) สู่ระดับมาก (หัวลูกศรด้านบน) จากตัวอย่างข้างต้นผลสัมฤทธิ์การเรีของนักเรียนแบ่งออกเป็นสี่ระดับก็สามารถนำคุณลักษณะของนักเรียนที่สอดคล้องกับระดับผลสัมฤทธิ์ดังกล่าวมาระบุตำแหน่งบนเส้นสมมตินี้ได้พร้อมกับมีคำบรรยายคุณลักษณะประกอบด้วยว่า นักเรียนที่มีความสามารถตรงกับระดับนั้นๆ รู้อะไรหรือมีทักษะอะไรหรือทำอะไรได้

แนวคิดนี้เป็นการสร้างมโนทัศน์เกี่ยวกับมาตรฐานความสามารถของนักเรียน จากนั้นเมื่อสร้างแบบสอบเพื่อประเมินความสามารถของนักเรียน เส้นตรงที่อยู่ทางขวาจะถูกสร้างขึ้นเพื่อบอก

สเกลคะแนนรวมจากแบบสอบนั้น ข้อสังเกตของเส้นตรงทางขวามือคือ ไม่มีหัวลูกศรทั้งสองข้าง เนื่องจากทราบพิสัยของคะแนนว่าเริ่มต้นและสิ้นสุดที่ค่าเท่าไร สเกลนี้เกิดจากค่าที่สังเกตได้ ในขณะที่เส้นตรงทางซ้ายมือเกิดจากข้อสันนิษฐานทางทฤษฎี เมื่อนำคะแนนของนักเรียนมาระบุตำแหน่งลงบนเส้นทางด้านขวาก็จะเห็นความเชื่อมโยงกับเส้นมาตรฐานความสามารถทางด้านซ้าย สมมติว่ามีนักเรียนทำแบบสอบจำนวนเจ็ดคนคือ นักเรียน A, B, C, D, E, F, และ G คะแนนของทุกคนจะถูกระบุตำแหน่งลงบนเส้นคะแนนทางด้านขวา การแจกแจงคะแนนสอบของนักเรียน ในลักษณะนี้จะให้แนวทางการกำหนดคะแนนจุดตัดว่าควรมีค่าเท่าไร ทำให้คะแนนจุดตัดที่เกิดขึ้น เป็นค่าที่สามารถสังเกตได้และมีลักษณะเป็นนิยามเชิงปฏิบัติการของระดับความสามารถที่คาดหวัง พร้อมทั้งมีคำบรรยายความสามารถของนักเรียนแต่ละระดับด้วย

ข้อสังเกตอีกประการหนึ่งเกี่ยวกับนิยามเชิงปฏิบัติการคือ แนวคิดดั้งเดิมที่พยายามบรรยายว่าผลสัมฤทธิ์ทางการเรียนคืออะไรและประกอบด้วยอะไรบ้างนั้นให้ข้อมูลไม่เพียงพอสำหรับสนับสนุนหรือโต้แย้งเกี่ยวกับความสามารถนักเรียน นิยามเชิงปฏิบัติการแบบนี้ให้ข้อมูลเพียงคร่าวๆ เกี่ยวกับโครงสร้างทางทฤษฎีของสิ่งที่วัด ทั้งๆ ที่นิยามเชิงปฏิบัติการ นอกจากจะต้องให้สารสนเทศว่าสิ่งที่วัดคืออะไร ประกอบด้วยอะไรบ้างแล้ว ยังจะต้องบอกระดับหรือปริมาณของสิ่งที่มุ่งวัดซึ่งสะท้อนผลสัมฤทธิ์ทางการเรียนที่แตกต่างกัน ดังนั้นนิยามเชิงปฏิบัติการของผลสัมฤทธิ์ทางการเรียนอีกแนวคิดหนึ่งจึงเน้นไปที่การบรรยายความสามารถของนักเรียนในแต่ละระดับ เพื่อให้ผลการวัดมีรายละเอียดมากขึ้นสามารถนำไปใช้สรุปอ้างอิงความสามารถของนักเรียนได้ถูกต้องและน่าเชื่อถือมากยิ่งขึ้นด้วย

ธรรมชาติของกระบวนการกำหนดคะแนนจุดตัดมีสามประการคือ ประการแรก คะแนนจุดตัดอาจจะกำหนดไว้ก่อนทราบคะแนนสอบของนักเรียนก็ได้ โดยผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจจะเป็นผู้พิจารณาระดับความสามารถที่คาดหวัง เช่น คะแนนขั้นต่ำที่นักเรียนจะ (would) ทำได้หรือจำนวนข้อสอบที่นักเรียนจะ (would) ตอบถูก ไม่ใช่คะแนนขั้นต่ำที่นักเรียนควรจะ (should) ทำได้หรือจำนวนข้อสอบที่นักเรียนควรจะ (should) ตอบถูก (Cizek, 2012) คะแนนจุดตัดในลักษณะนี้จะไม่แน่นอนและขาดหลักฐานที่น่าเชื่อถือมาสนับสนุนแนวคิดของผู้มีอำนาจตัดสินใจว่าใช้แนวคิดใดกำหนดคะแนนจุดตัด (Perie, 2008) ในบางสถานการณ์คะแนนจุดตัดอาจจะกำหนดไว้อย่างคร่าวๆ จากนั้นผู้เชี่ยวชาญจะอาศัยผลวิเคราะห์คะแนนสอบช่วยยืนยันความถูกต้องและเหมาะสมอีกครั้ง ประการที่สอง การกำหนดคะแนนจุดตัดไม่จำเป็นต้องแบ่งสเกลคะแนนออกเป็นช่วงที่เท่ากัน ตัวอย่างเช่น จากภาพที่ 2.1 ขั้นผู้เชี่ยวชาญอาจจะมิพิสัยของคะแนนที่กว้างกว่าขั้นอื่นๆ ก็ได้ ทั้งนี้คะแนนจุดตัดควรสัมพันธ์กับระดับความสามารถที่คาดหวังเพื่อจำแนกนักเรียนออกเป็นกลุ่มๆ ได้อย่างมีประสิทธิภาพ ประการสุดท้ายคือ กระบวนการกำหนดคะแนนจุดตัดต้องยุติธรรมกับนักเรียนทุกคน ไม่ทำให้นักเรียนกลุ่มใดได้ประโยชน์มากกว่ากลุ่มอื่น อีกทั้งคะแนนจุดตัดจะต้องมีความตรงเชิงตีความ (interpretive validity) (Cizek, 2012; Kane, 1994; Mehrens & Cizek, 2012; Wyse, 2013) คือสามารถตีความหมายคะแนนที่อยู่ต่ำกว่าและสูงกว่าคะแนนจุดตัดได้ คะแนนจุดตัดจะต้อง

สะท้อนระดับความสามารถที่เป็นจริงของนักเรียนแต่ละกลุ่ม และผลจากการใช้คะแนนจุดตัดสามารถใช้สรุปอ้างอิงคุณลักษณะของนักเรียนและนำผลที่ได้ไปใช้ตรงตามวัตถุประสงค์ที่กำหนดไว้

คุณลักษณะของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์

การกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ (Construct Mapping method: Draney & Kennedy, 2010; Draney, Kennedy, Moore, & Morell, 2010; Draney & Wilson, 2011; Wilson & Draney, 2002; Wyse, 2013) ถูกพัฒนาขึ้นภายใต้กรอบการพัฒนาการประเมินที่เรียกว่ากรอบการสร้างโมเดลภาวะสันนิษฐาน (Construct Modeling framework: Wilson, 2005) ภายในกรอบแนวคิดนี้จะมีแผนที่ภาวะสันนิษฐาน (construct map) ทำหน้าที่เป็นแกนเชื่อมโยงส่วนประกอบต่างๆ ภายในกระบวนการประเมิน ดังนั้นเมื่อนำแนวคิดของแผนที่ภาวะสันนิษฐานมาใช้ในกระบวนการกำหนดคะแนนจุดตัด แผนที่ภาวะสันนิษฐานจะทำหน้าที่เชื่อมโยงส่วนประกอบต่างๆ ของกระบวนการทดสอบเข้ากับกระบวนการกำหนดคะแนนจุดตัด

เนื่องจากชื่อภาษาอังกฤษของทั้งสามคำคล้ายคลึงกัน ผู้วิจัยจึงขอใช้คำภาษาไทยโดยพิจารณาจากเนื้อหาสาระของแต่ละคำเพื่อป้องกันความสับสนและความเข้าใจคลาดเคลื่อน กล่าวคือ construct modeling framework เป็นกรอบแนวคิดหรือเป็นตรรกะที่แสดงถึงขั้นตอนและส่วนประกอบต่างๆ ในกระบวนการพัฒนาการประเมินหรือการพัฒนาการทดสอบที่ให้ความสำคัญกับการสร้างโมเดลหรือการออกแบบโครงสร้างทางทฤษฎีของสิ่งที่นักวิจัยต้องการวัด โดยสิ่งที่ต้องการวัดมีลักษณะเป็นตัวแปรแฝง โครงสร้างภายในของตัวแปรแฝงนี้จะมีรายละเอียดตามการสันนิษฐานของผู้วิจัยจากสิ่งที่สังเกตได้ ดังนั้นความหมายของคำว่า construct จึงไม่ได้สะท้อนเพียงแค่ส่วนประกอบของ construct นั้น แต่แสดงถึงโครงสร้างและรายละเอียดของ construct ด้วย เช่นเดียวกับโครงสร้างของบ้านที่มีทั้งส่วนประกอบของโครงสร้างบ้านและพื้นที่ส่วนต่างๆ ภายในบ้านที่บ่งบอกหน้าที่การใช้งานแตกต่างกัน construct จึงมีลักษณะเป็นภาวะสันนิษฐานที่ผู้วิจัยคาดว่าจะเป็นอย่างนั้นในตอนแรก แล้วจะชัดเจนขึ้นเมื่อนำภาวะสันนิษฐานนั้นไปใช้ ดังนั้นคำว่า Construct Modeling framework จึงน่าจะเหมาะสมกับคำภาษาไทยว่า กรอบการสร้างโมเดลภาวะสันนิษฐาน

ภายในกรอบการสร้างโมเดลภาวะสันนิษฐานจะมีแผนที่ภาวะสันนิษฐานเป็นหนึ่งในองค์ประกอบหลักที่สำคัญที่สุด (Wilson, 2005; Wilson et al., 2012) เนื่องจากแผนที่ภาวะสันนิษฐานทำหน้าที่เป็นโมเดลให้กับสิ่งที่ต้องการวัดหรือเป็นโครงสร้างของ construct โดยจะนำรายละเอียดของ construct มาระบุตำแหน่งบนเส้นสมมติแสดงปริมาณการมีคุณลักษณะจากน้อยไปสู่มาก แนวคิดดังกล่าวจึงใกล้เคียงกับการทำแผนที่ของสถานที่ต่างๆ สำหรับแผนที่ภาวะสันนิษฐานจึงเป็นความเชื่อที่นักวิจัยระบุตำแหน่งคร่าวๆ ของรายละเอียดต่างๆ บนแผนที่ที่สมมติขึ้นตามทฤษฎี ดังนั้น construct map จึงน่าที่จะมีชื่อเรียกภาษาไทยว่า แผนที่ภาวะสันนิษฐาน และเรียก construct mapping ว่าการทำหรือการสร้างแผนที่ภาวะสันนิษฐาน

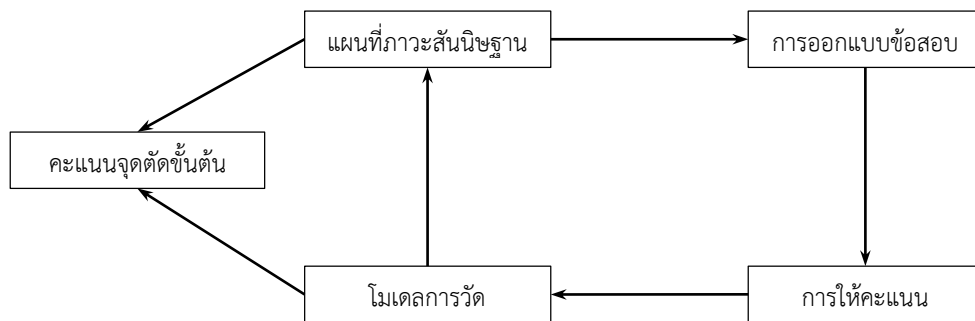
จากที่ได้อธิบายไปแล้ว แผนทีภาวะสันนิษฐานเกิดขึ้นตามความเชื่อของนักวิจัยที่ได้ค้นคว้าและวิเคราะห์ทฤษฎีที่เกี่ยวข้องหรือแม้กระทั่งจากประสบการณ์และข้อมูลต่างๆ เมื่อแผนที่ภาวะสันนิษฐานถูกนำไปใช้สร้างข้อสอบหรือข้อคำถามก็จะถูกตรวจสอบความถูกต้องด้วยทฤษฎีการวัดของราสช์ (Rasch measurement theory: Rasch, 1960; Wright & Stone, 1979) ผลวิเคราะห์ที่ได้จะถูกนำไปใช้ปรับปรุงแผนที่ภาวะสันนิษฐานที่ได้ออกแบบไว้ตั้งแต่ต้น ดังนั้นแผนที่ภาวะสันนิษฐานในตอนแรกจึงเป็นแผนที่เชิงทฤษฎี (theoretical construct map: Wyse, 2013) ส่วนแผนที่ที่เกิดขึ้นภายหลังการปรับปรุงแก้ไขแล้วจะเป็นแผนที่เชิงประจักษ์ (empirical construct map: Wyse, 2013) ดังนั้นจึงไม่ควรใช้ชื่อภาษาไทยเรียกแทน construct map ว่าแผนที่ภาวะสันนิษฐานเชิงทฤษฎี เพราะแผนที่ดังกล่าวเป็นได้ทั้งแผนที่เชิงทฤษฎีและแผนที่เชิงประจักษ์ หรือเรียกว่า แผนที่เชิงโครงสร้าง เพราะคำว่า แผนที่ และ โครงสร้าง มีสื่อความหมายเดียวกัน

ในกระบวนการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ เป็นการนำแผนที่ภาวะสันนิษฐานเชิงประจักษ์มาใช้กำหนดค่าที่จะสามารถเป็นคะแนนจุดตัดซึ่งเส้นสมมติที่ใช้แสดงปริมาณการมีคุณลักษณะของสิ่งที่สนใจและเป็นเส้นบอกตำแหน่งของบุคคลที่จะมีคุณลักษณะนั้นๆ ในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี จะถูกเปลี่ยนเป็นสเกลความสามารถแฝงหรือสเกลค่าประมาณความสามารถแฝง θ ที่ได้จากการวิเคราะห์ด้วยโมเดลการวัดของราสช์ดังนั้นชื่อภาษาไทยที่ใช้เรียกวิธีกำหนดคะแนนจุดตัด Construct Mapping method จึงน่าจะเรียกว่า วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ เพราะภาวะสันนิษฐานถูกแปลงเป็นสเกลความสามารถแฝงที่ได้มาจากการวิเคราะห์ข้อมูลเชิงประจักษ์

ตามที่คุณวิจัยได้เน้นย้ำว่า การกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์เป็นการเชื่อมโยงกระบวนการกำหนดคะแนนจุดตัดเข้ากับกระบวนการพัฒนาการทดสอบ การพัฒนาระบบการทดสอบภายใต้กรอบการสร้างโมเดลภาวะสันนิษฐานจะมีแผนที่ภาวะสันนิษฐานเป็นองค์ประกอบที่สำคัญที่สุดและถือว่าเป็นจุดเริ่มต้นของขั้นตอนต่างๆ ที่จะเกิดขึ้นต่อเนื่องกันไปจนสิ้นสุดกระบวนการพัฒนาการทดสอบ คุณลักษณะเด่นที่สำคัญของกรอบการสร้างโมเดลภาวะสันนิษฐานคือ จะเชื่อมโยงมโนทัศน์ของโมเดลทักษะการคิด (cognition model) ซึ่งมักจะถูกกละเลยในการวัดทางการศึกษาและจิตวิทยา (Mislevy, 1996; National Research Council, 2001) เข้ากับการออกแบบข้อสอบ (item design) แนวทางการให้คะแนนคำตอบ (scoring guide) และโมเดลการวัดเพื่อวิเคราะห์คำตอบ (measurement model) นั้นหมายความว่าผลวิเคราะห์เกี่ยวกับคุณลักษณะข้อสอบ (item properties) และคุณลักษณะผู้สอบ (examinee properties) จะถูกติดตามตามโมเดลทักษะการคิดที่ได้ออกแบบไว้ตั้งแต่ขั้นตอนแรก (Brown & Wilson, 2011; Wilson, 2005; Wilson et al., 2012)

เมื่อนำกรอบการสร้างโมเดลภาวะสันนิษฐาน โดยเฉพาะอย่างยิ่งแผนที่ภาวะสันนิษฐานมาประยุกต์ใช้กำหนดคะแนนจุดตัดดังภาพที่ 2.2 จะเห็นว่าเป็นวิธีการที่นำสารสนเทศที่เกี่ยวข้องกับโครงสร้างภายใน (internal structure) ของตัวแปรมาสนับสนุนความตรงเชิงโครงสร้าง (construct

validity) ของคะแนนจุดตัด เพราะคะแนนจุดตัดจะถูกกำหนดตามโครงสร้างของสิ่งที่ต้องการวัด มีหลักฐานสนับสนุนความตรงเชิงตีความ (interpretative validity) เพราะคะแนนจุดตัดจะตีความเป็นความสามารถของนักเรียนตามที่กำหนดไว้ในโครงสร้างของสิ่งที่ต้องการวัด และมีหลักฐานสนับสนุนความตรงตามจุดมุ่งหมายหรือความตรงของผลลัพธ์ที่เกิดขึ้น (consequential validity) เพราะคะแนนจุดตัดซึ่งกำหนดมาจากคะแนนสอบจะถูกนำไปใช้ตรงตามวัตถุประสงค์ของการทดสอบ



ภาพที่ 2.2 การประยุกต์กรอบการสร้างโมเดลภาวะสันนิษฐานเพื่อกำหนดคะแนนจุดตัด

ข้อสังเกตเกี่ยวกับการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ คือ แนวคิดเริ่มแรกของความสามารถของนักเรียนจะถูกพิจารณาเป็นค่าต่อเนื่อง (continuous scale) แล้วจึงถูกแปลงเป็นแบบจัดอันดับ (ordinal scale) ทำให้คุณสมบัติของสเกลตัวแปรแฝงบางอย่างสูญหายไป แต่ก็ช่วยให้การตัดสินความสามารถของนักเรียนว่าใครควรผ่านหรือไม่ผ่านเกณฑ์การประเมินมีความเป็นปรนัยมากขึ้น และช่วยสื่อสารระหว่างผู้เชี่ยวชาญกับผู้มีส่วนได้เสียให้เข้าใจผลวิเคราะห์ข้อมูลและสารสนเทศต่างๆ มากขึ้น (Kane & Tannenbaum, 2013)

แผนที่ภาวะสันนิษฐานและคำบรรยายระดับความสามารถ

คำบรรยายระดับความสามารถ (performance level descriptor) เป็นสิ่งจำเป็นอย่างมากในกระบวนการกำหนดคะแนนจุดตัด (ศิริพันธ์ ดิยะวงศ์สุวรรณ ฌัญญุภรณ์ หลาวทอง และ ศิริชัย กาญจนวาสี, 2556; Cizek & Bunch, 2007; Egan, Schnieder, & Ferrara, 2012; Perie, 2008; Wyse, 2013, 2014; Zieky, 2012) เพราะคำบรรยายระดับความสามารถจะถูกนำไปใช้ตีความคะแนนสอบและคะแนนจุดตัดเช่นเดียวกับแผนที่ภาวะสันนิษฐานที่สำคัญอย่างมากในกระบวนการทดสอบ คำบรรยายระดับความสามารถเป็นชุดคำบรรยายความรู้ (knowledge) ทักษะ (skill) และความสามารถ (ability) ของนักเรียนในแต่ละระดับความสามารถ (Cizek & Bunch, 2007; Perie, 2008) คำบรรยายระดับความสามารถมีลักษณะเป็นระดับขั้นความสามารถจากต่ำสุดสู่สูงสุดภายใต้ขอบเขตเฉพาะ ระดับความสามารถที่สูงขึ้นไปเรื่อยๆ จะแสดงถึงการมีความรู้ ทักษะและความสามารถเพิ่มขึ้นไปเรื่อยๆ ตามเส้นต่อเนื่องผลสัมฤทธิ์ (Plake, Huff, & Reshetar, 2010)

ภาพที่ 2.3 แสดงตัวอย่างคำบรรยายระดับความสามารถทางวิทยาศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 5 ของรัฐ Nebraska และภาพด้านล่างเป็นคำบรรยายระดับความสามารถด้านการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 4 ของรัฐ Texas ที่ใช้ในกระบวนการกำหนดคะแนนจุดตัดและตีความคะแนนสอบของนักเรียน ส่วนภาพที่ 2.4 และ 2.5 เป็นตัวอย่างแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์และทักษะการสื่อสารซึ่งถูกนำไปใช้ในกระบวนการสร้างแบบสอบและออกแบบการประเมิน จากตัวอย่างข้างต้นจะเห็นว่าคำบรรยายระดับความสามารถและแผนที่ภาวะสันนิษฐานมีลักษณะและรายละเอียดคล้ายคลึงกันหลายประการ ไม่ว่าจะเป็นการเขียนคำบรรยายความสามารถในแต่ละระดับ มุมมองที่มีต่อความสามารถนักเรียนแบบค่าต่อเนื่องที่เพิ่มขึ้นจากระดับต่ำสู่ระดับที่สูงขึ้น และการออกแบบระดับความสามารถเป็นลำดับขั้น

แต่อย่างไรก็ตามแผนที่ภาวะสันนิษฐานมีคุณลักษณะเด่นกว่าคำบรรยายระดับความสามารถสามประการ กล่าวคือ ประการแรก แผนที่ภาวะสันนิษฐานสามารถใช้พัฒนาข้อสอบและแบบสอบแต่ในขณะที่คำบรรยายระดับความสามารถอาจจะใช้ไม่ได้ เนื่องจากคำบรรยายระดับความสามารถไม่มีคำบรรยายคำตอบของนักเรียนในแต่ละระดับ คำตอบของนักเรียนจะเป็นหลักฐานอย่างดีเพื่อตีความและสรุปอ้างอิงความสามารถของนักเรียนให้ชัดเจนขึ้น (Brown & Wilson, 2011; Wilson, 2005; Wilson et al., 2012)

ประการที่สอง คำบรรยายระดับความสามารถสนใจคะแนนรวมจากแบบสอบ แต่แผนที่ภาวะสันนิษฐานสนใจทั้งคะแนนรวมและคะแนนรายข้อ ดังนั้นเมื่อนำแผนที่ภาวะสันนิษฐานไปใช้เป็นคำบรรยายระดับความสามารถในกระบวนการกำหนดคะแนนจุดตัด ซึ่งนอกจากจะได้สารสนเทศละเอียดมากขึ้นแล้ว ยังเชื่อมโยงความหมายของคะแนนรายข้อกับความหมายของคะแนนรวม คำตอบของนักเรียนที่มีต่อข้อสอบแต่ละข้อจึงสามารถใช้เป็นหลักฐานสนับสนุนความตรงเชิงโครงสร้าง (Luecht, 2013; Wilson, 2005) และความตรงเชิงตีความ (Kane, 2006) ได้

ประการที่สาม แผนที่ภาวะสันนิษฐานสามารถใช้เป็นฐานการพัฒนาคำบรรยายระดับความสามารถในกรณีที่ไม่มีคำบรรยายระดับความสามารถในกระบวนการกำหนดคะแนนจุดตัด (Schulz, 2013) และในกรณีที่การทดสอบแบบดั้งเดิมไม่ได้กำหนดคำบรรยายระดับความสามารถไว้ก่อน มีเพียงแต่ข้อกำหนดลักษณะแบบสอบ (test specification) (Luecht, 2013; Masters, 2010) ซึ่ง Schulz (2013) และ Wyse (2013, 2014) บ่งชี้ว่าแผนที่ภาวะสันนิษฐานมีส่วนช่วยในกระบวนการพิจารณาความเหมาะสมของคะแนนจุดตัด และช่วยเชื่อมโยงข้อมูลที่เกี่ยวข้องกับคะแนนสอบเข้ากับคำบรรยายระดับความสามารถ ในกรณีที่ผู้เชี่ยวชาญมีความเห็นไม่ตรงกัน แผนที่ภาวะสันนิษฐานจะช่วยให้ผู้เชี่ยวชาญแต่ละคนเข้าใจความคิดของผู้เชี่ยวชาญคนอื่นๆ ด้วย

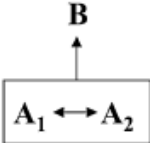
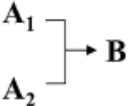
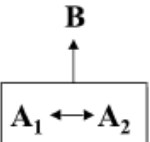
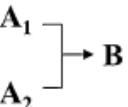
Nebraska State Accountability-Science (NeSA-S) Performance Level Descriptors Grade 11		
Below the Standards	Meets the Standards	Exceeds the Standards
Overall student performance in science reflects <i>unsatisfactory</i> performance on the standards and <i>insufficient</i> understanding of the content at eleventh grade. A student scoring at the Below the Standards level <i>inconsistently</i> draws on a broad range of scientific knowledge and skills in the areas of inquiry, physical, life, and Earth/space sciences. A student at this level <i>inconsistently</i>: - Identifies a testable hypothesis that guides a scientific investigation. - Identifies and manages variables and constraints. - Identifies appropriate selection and use of scientific equipment. - Analyzes and interprets data and evaluates models and explanations. - Evaluates scientific investigations. - Identifies matter in terms of its structure and composition. - Identifies field forces. - Identifies energy systems in matter. - Identifies organic molecules, sub-cellular structures, and cellular functions. - Identifies DNA and its role in heredity. - Identifies the flow of energy between organisms and their environment. - Identifies types of adaptations necessary for survival. - Identifies components of the universe. - Identifies the relationships between Earth's structure and processes. - Identifies relationships between sources of energy and Earth's systems. - Identifies the Law of Superposition.	Overall student performance in science reflects <i>satisfactory</i> performance on the standards and <i>sufficient</i> understanding of the content at eleventh grade. A student scoring at the Meets the Standards level <i>generally</i> draws on a broad range of scientific knowledge and skills in the areas of inquiry, physical, life, and Earth/space sciences. A student at this level <i>generally</i>: - Identifies a testable hypothesis that guides a scientific investigation. - Identifies and manages variables and constraints. - Identifies appropriate selection and use of scientific equipment. - Analyzes and interprets data and evaluates models and explanations. - Evaluates scientific investigations. - Describes matter in terms of its structure, composition, and conservation. - Describes the nature of field forces and their interactions with matter. - Describes energy systems relating to the conservation and interaction of energy and matter. - Describes the chemical basis of the growth, development, and maintenance of cells. - Describes the molecular basis of reproduction and heredity. - Describes, on a molecular level, the cycling of matter and the flow of energy between organisms and their environment. - Describes the theory of biological evolution. - Describes the known universe. - Investigates the relationships among Earth's structure, systems, and processes. - Describes the relationships among the sources of energy and their effects on Earth's systems. - Explains the history and evolution of Earth.	Overall student performance in science reflects <i>high academic</i> performance on the standards and a <i>thorough</i> understanding of the content at eleventh grade. A student scoring at the Exceeds the Standards level <i>consistently</i> draws on a broad range of scientific knowledge and skills in the areas of inquiry, physical, life, and Earth/space sciences. A student at this level <i>consistently</i>: - Identifies a testable hypothesis that guides a scientific investigation. - Identifies and manages variables and constraints. - Identifies appropriate selection and use of scientific equipment. - Analyzes and interprets data and evaluates models and explanations. - Evaluates scientific investigations. - Analyzes structure, composition, and conservation of matter. - Analyzes interactions between field forces and matter. - Analyzes interactions between energy systems and matter. - Analyzes the chemical basis of the growth, development, and maintenance of cells. - Analyzes the molecular basis of reproduction and heredity. - Analyzes the cycling of matter and the flow of energy between organisms and their environment. - Analyzes the theory of biological evolution and the diversity of life. - Analyzes the formation of the universe. - Analyzes the relationships among Earth's structure, systems, and processes. - Analyzes the relationships between sources of energy and their effects on Earth's systems. - Analyzes the history and evolution of Earth.

ที่มา: http://www.education.ne.gov/assessment/pdfs/NeSA_Science_PLDs.pdf, 2559

Performance Level Descriptors
Students achieving Level III: Advanced Academic Performance can - Write thoughtful and engaging personal narratives that skillfully convey the experience by including details and vivid, expressive language that strengthen the cohesiveness of the narrative, significantly contributing to the writer's portrayal of the experience - Write thoughtful and engaging expository essays with sustained focus on a clear central idea, an organizing structure that is well suited to the expository task, and development that adds substance to the essay - Choose sentences that are purposeful, varied, and well controlled
Students achieving Level II: Satisfactory Academic Performance can - Write credible personal narratives that are focused on a specific personal experience, use organizational strategies or literary devices that contribute to narrative writing, include specific details and language that support a portrayal of the experience, and demonstrate an adequate command of age-appropriate written conventions - Write expository essays that establish a clear central idea; use organizational strategies that are adequately suited to expository writing; sufficiently develop ideas with specific details, language, and examples that are appropriate to the explanatory purpose; and demonstrate an adequate command of age-appropriate written conventions - Use correct sentence structures - Revise drafts to strengthen the development of ideas by adding or deleting information, improve the sentence-to-sentence connections within paragraphs, and ensure that word choice is appropriate and sentences are controlled - Edit drafts to correct errors in grammar, sentence structure, capitalization, punctuation, and spelling
Students achieving Level I: Unsatisfactory Academic Performance can - Write basic or limited personal narratives that demonstrate a minimal understanding of the writing task and convey some sense of the experience - Write basic or limited essays that are marginally suited to the expository purpose and minimally developed - Demonstrate a partial command of written conventions and sentence structures - Demonstrate limited skills in revision and editing

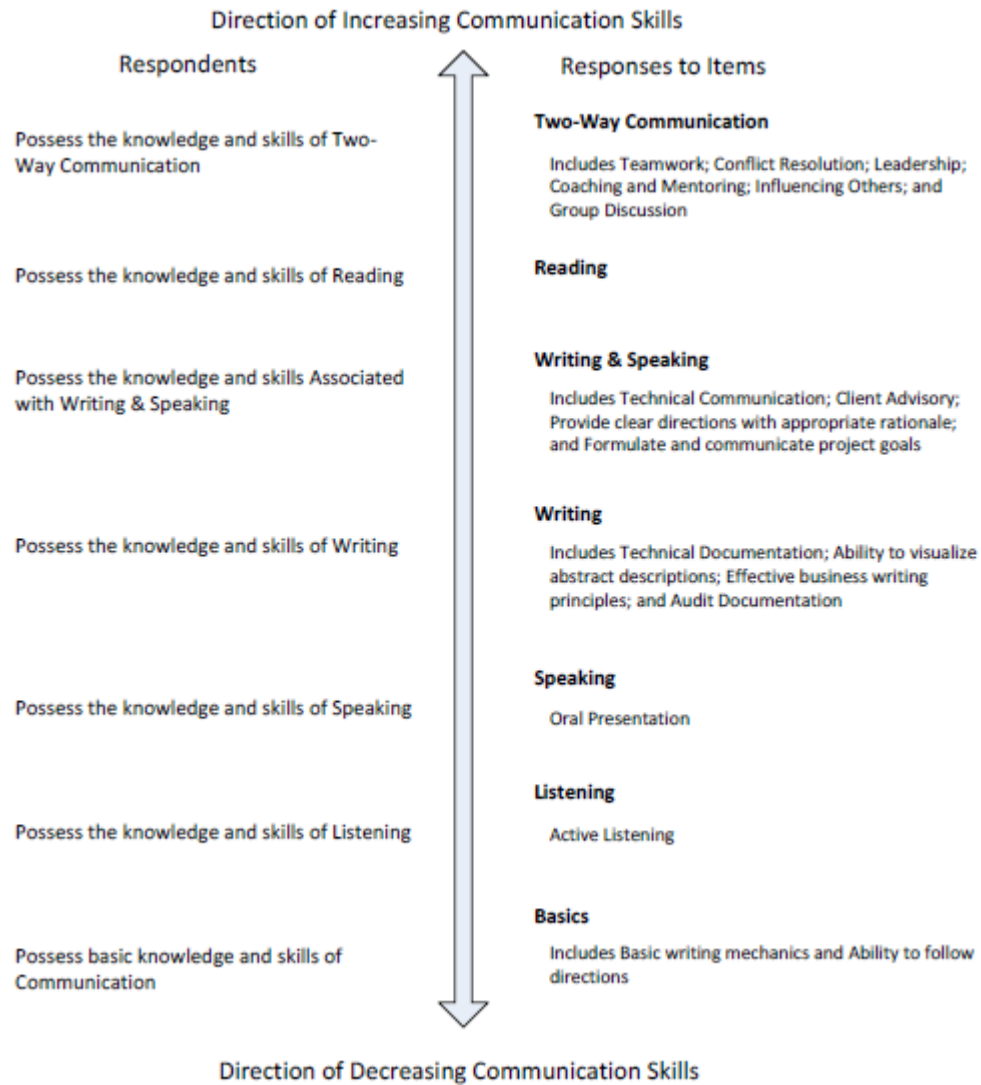
ที่มา: http://www.ecisd.org/sites/ecisd.org/files/4th_grade_writing_descriptors.pdf, 2559

ภาพที่ 2.3 ตัวอย่างคำบรรยายระดับความสามารถทางวิทยาศาสตร์และทักษะการเขียน

Description of Person		
Emergent		The phenomenon is seen as an emergent property of a system, made up of interacting components. The system evolves over time, eventually producing the observed effect.
Multiple		The phenomenon is seen as an effect produced by multiple causal elements. All are necessary; if one is removed the effect is not produced.
Justified		The phenomenon is seen as an effect produced by a single causal element. Justification or a mechanism is necessary.
Causal		The phenomenon is seen as an effect produced by a single causal element. Justification or a mechanism is not necessary.
Acausal	B	The phenomenon is seen as an instantiation of reality. No cause is necessary.
None	B !?	The phenomenon is surprising. No explanation seems possible.
Description of Response		
		Score
Emergent		"A ₁ and A ₂ happen. Over time, they interact and evolve, until eventually B happens. Meanwhile, A ₁ and A ₂ continue to happen." 5
Multiple		"A ₁ and A ₂ cause B when they both happen at the same time." 4
Justified		"A causes B, and this is how." 3
Causal		"A causes B." 2
Acausal	B	"B happens because that's what happens." 1
None	B !?	"I can't explain why B happens." 0

ที่มา: A model of cognition: The missing cornerstone of assessment (Brown & Wilson, 2011)

ภาพที่ 2.4 ตัวอย่างแผนที่ภาวะสันนิษฐานการให้เหตุผลทางวิทยาศาสตร์



ที่มา: A review of assessment engineering principles with select application to the certified public accountant examination: Technical report (Zhou, 2009)

ภาพที่ 2.5 ตัวอย่างแผนที่ภาวะสันนิษฐานทักษะการสื่อสาร

จุดเด่นของวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์

วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์มีคุณสมบัติเด่นสี่ประการ คือ ประการแรก วิธีดังกล่าวสามารถใช้กำหนดคะแนนจุดตัดของคะแนนสอบที่มีจากแบบสอบที่เป็นเอกมิติและพหุมิติ กล่าวคือวิธีดังกล่าวจะไม่ฝ่าฝืนธรรมชาติของแบบสอบและธรรมชาติของความสามารถนักเรียนที่ใช้ทำข้อสอบ เนื่องจากการทดสอบส่วนใหญ่จะเกี่ยวข้องกับสิ่งที่ต้องการวัดหลายด้าน เช่น การประเมินความพร้อมเพื่อเรียนต่อในระดับอุดมศึกษาที่แบบสอบหนึ่งฉบับอาจจะประกอบด้วยข้อสอบที่วัดทั้งความสามารถทางการคำนวณและความสามารถทางภาษา ความสามารถนักเรียนที่ใช้ทำแบบสอบ

ฉบับนี้จึงประกอบด้วยหลายด้าน ดังนั้นการกำหนดคะแนนจุดตัดจึงควรให้ความสำคัญกับความเป็นพหุมิติ (multidimensionality) ของแบบสอบ โดยในขั้นตอนวิเคราะห์คำตอบ ค่าประมาณความสามารถนักเรียนสามารถจำแนกเป็นรายด้าน แทนที่จะมีเพียงค่าเดียวตามข้อตกลงดั้งเดิมเกี่ยวกับความเป็นเอกมิติ (unidimensionality) ของแบบสอบตามแบบวิธีการกำหนดคะแนนจุดตัดแบบดั้งเดิม เช่นวิธีของ Angoff (1971) วิธีของ Nedelsky (1954) วิธีของ Ebel (1972) และวิธีของ Jaeger (1982)

แม้ว่าการกำหนดคะแนนจุดตัดข้างต้นจะประยุกต์ใช้ได้ในกรณีที่แบบสอบประกอบด้วยเนื้อหาสาระหลายส่วน โดยแยกคะแนนสอบของนักเรียนเป็นรายด้านแล้วจึงกำหนดคะแนนจุดตัดของแต่ละด้าน แต่อย่างไรก็ตามแนวทางดังกล่าวยังคงเป็นแนวคิดแบบเอกมิติเหมือนเดิมหรือที่เรียกว่า consecutive unidimensional approach (Briggs & Wilson, 2003) โดยไม่นำความสัมพันธ์ระหว่างความสามารถด้านต่างๆ มาใช้ประมาณค่าความสามารถของนักเรียนแล้วจึงกำหนดคะแนนจุดตัด นอกจากนี้แม้ว่าวิธีของ Plake และ Hambleton (2001) วิธีของ Jaeger และ Mills (2001) วิธีของ Putnam, Pence และ Jaeger (1995) วิธีของ Jaeger (1995) และวิธีของ Kingston และ Tiemann (2012) ให้ความสำคัญกับความเป็นพหุมิติของแบบสอบ แต่ไม่เหมาะสมเมื่อมีข้อสอบบางข้อวัดมากกว่า 1 เนื้อหาย่อย (within-item multidimensionality) (Draney & Wilson, 2011; Kingston, Tiemann, & Loughran, 2013; Wyse, 2013) วิธีเหล่านี้จึงเหมาะสมกับข้อสอบแต่ละข้อที่วัดเพียงเนื้อหาย่อยเดียว (between-item multidimensionality)

ประการที่สอง วิธีดังกล่าวเชื่อมโยงคำบรรยายระดับความสามารถเข้ากับค่าที่จะใช้เป็นคะแนนจุดตัดได้อย่างกลมกลืน กล่าวคือกระบวนการกำหนดคะแนนจุดตัดควรเริ่มต้นด้วยการพัฒนาคำบรรยายระดับความสามารถของนักเรียนว่า ความรู้และทักษะที่นักเรียนในแต่ละระดับความสามารถจะทำได้คืออะไร (Cizek & Bunch, 2007; Perie, 2008) อย่างไรก็ตาม สำหรับวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์นั้น ไม่จำเป็นที่จะต้องสร้างคำบรรยายระดับความสามารถขึ้นมาก่อนก็ได้ (Schulz, 2013) เพราะจะมีแผนที่ภาวะสันนิษฐานในกระบวนการพัฒนาการทดสอบที่ทำหน้าที่แทนคำบรรยายระดับความสามารถในกระบวนการกำหนดคะแนนจุดตัด แผนที่ภาวะสันนิษฐานประกอบด้วยโครงสร้างความสามารถสมมติเบื้องต้นที่ผนวกรวมคำบรรยายระดับความสามารถและคำบรรยายคำตอบที่สอดคล้องกับระดับความสามารถนั้นๆ เข้าไว้ด้วยกัน จากนั้นจึงสร้างข้อสอบตามแผนที่ภาวะสันนิษฐาน ให้คะแนนคำตอบและประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบ ทำให้พารามิเตอร์ผู้สอบที่ได้เชื่อมโยงเป็นอย่างดีกับคำบรรยายระดับความสามารถ ช่วยให้การตัดสินใจคะแนนจุดตัดมีความสมเหตุสมผลบนพื้นฐานของข้อมูลเชิงประจักษ์ (Bunch, 2013; Kane & Tannenbaum, 2013; McClarty, 2013; Skaggs, 2013)

เมื่อนำคุณลักษณะประการที่สองเชื่อมโยงเข้ากับประการที่หนึ่งพบว่า กระบวนการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ไม่จำเป็นต้องมีคำบรรยายระดับความสามารถเพียงด้านเดียว สามารถมีคำบรรยายระดับความสามารถรายด้านตามเนื้อหาที่มุ่งวัด

และคำบรรยายระดับความสามารถย่อยก็สามารถเชื่อมต่อกันด้วยโมเดลวิเคราะห์ความสามารถแฝงของนักเรียนทำให้คะแนนจุดตัดแต่ละด้านสามารถตีความเชื่อมโยงกันได้

ประการที่สาม วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์จัดระบบและสรุปรวมสารสนเทศที่จำเป็นต่อการกำหนดคะแนนจุดตัดให้อยู่ในรูปแผนภาพ เพราะเมื่อผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจเข้าใจความสัมพันธ์ระหว่างสารสนเทศต่างๆ มากขึ้น ก็จะช่วยให้การตัดสินใจเลือกค่าใดเป็นคะแนนจุดตัดมีความชัดเจนมากขึ้นด้วย (Bunch, 2013; Davis-Becker, 2013; Draney & Wilson, 2011; Kane & Tannenbaum, 2013; Skaggs, 2013; Wyse, 2013) วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์มีแผนที่ภาวะสันนิษฐานเป็นหัวใจสำคัญของการเชื่อมโยงสารสนเทศต่างๆ เข้าด้วยกัน เช่น พารามิเตอร์ข้อสอบและความสามารถผู้สอบ คะแนนดิบ รูปแบบข้อสอบและความคิดเห็นต่างๆ จากผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจคะแนนจุดตัด สารสนเทศที่เชื่อมโยงกันเหล่านี้จะทำหน้าที่เป็นคำอธิบายที่สมเหตุสมผลให้กับกระบวนการกำหนดคะแนนจุดตัดทั้งหมด (Kane & Tannenbaum, 2013; Schulz, 2013; Wyse, 2013)

อย่างไรก็ตามวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์แตกต่างและดีกว่าวิธีอื่นๆ ที่ใช้แผนที่ข้อสอบ (item map) เนื่องจากสามารถรวบรวมข้อมูลที่เกี่ยวข้องกับผู้สอบและข้อสอบมาใช้ประกอบการพิจารณาคะแนนจุดตัดได้ด้วย (Draney & Wilson, 2011) วิธีการกำหนดคะแนนจุดตัดในกลุ่มนี้เช่น วิธีของ Loomis และ Bourque (2001) วิธี Briefing Book (Haertel, 2002) วิธีของ Wang (2003) วิธี Bookmark (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012; Mitzel, Lewis, Patz, & Green, 2001) และวิธี Body of Work (Kingston, Kahl, Sweeney, & Bay, 2001; Kingston & Tiemann, 2012) ที่ให้ความสนใจอย่างมากกับโมเดลการวัดและความสัมพันธ์ระหว่างพารามิเตอร์ข้อสอบกับการแจกแจงพารามิเตอร์ผู้สอบ ทำให้ผู้เชี่ยวชาญหรือผู้มีอำนาจตัดสินใจขาดการเชื่อมโยงสารสนเทศที่จำเป็นอื่นๆ เข้ากับคำบรรยายระดับความสามารถไม่สามารถเชื่อมโยงความหมายของคะแนนจุดตัดเข้ากับความหมายของคะแนนสอบ และคะแนนจุดตัดมักไม่คงที่เมื่อแบบสอบมีข้อกำหนดลักษณะแบบสอบเปลี่ยนแปลงทุกปี (Draney & Wikson, 2011; Kane, 1994, 2001, 2006; Kane & Tannenbaum, 2013; Tannenbaum & Katz, 2013; Wyse, 2013)

ประการที่สี่ วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์เป็นฐานให้กับวิธีการกำหนดคะแนนจุดตัดอื่นๆ (Kane & Tannenbaum, 2013; Wyse, 2013; Wyse et al., 2013) กล่าวคือ โดยทั่วไปแล้วความก้าวหน้าของการกำหนดคะแนนจุดตัดตัดสินใจที่การพัฒนาวิธีวิเคราะห์ใหม่ๆ ซึ่งสามารถค้นคว้าได้จากงานวิจัยของ Cizek (2012), Cizek และ Bunch (2007), Kaftandjieva (2010), Hambleton และ Pitoniak (2006), Tannenbaum และ Katz (2013), Zieky, Perie, และ Livingston (2008) และ Wyse (2013) แต่อย่างไรก็ตาม วิธีเหล่านี้ดูเหมือนว่ายังไม่สามารถให้ข้อเสนอแนะแก่ผู้เชี่ยวชาญได้ว่าควรจะยอมรับ หรือปฏิเสธ หรือแก้ไขคะแนนจุดตัดอย่างไร (Kane & Tannenbaum, 2013; Schulz, 2013; Wyse, 2013) แต่สำหรับวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์นอกจากจะแสดง

ความสัมพันธ์ของสารสนเทศต่างๆ แล้วยังเสนอค่าที่สามารถเป็นคะแนนจุดตัดพร้อมกับมีข้อมูลประกอบการตัดสินใจเพื่อให้ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจได้เลือกค่าที่เหมาะสมหรือจะตัดแปลงค่าใดค่าหนึ่งให้เหมาะสมมากขึ้น แทนการกำหนดให้ใช้ค่าใดค่าเพียงค่าเดียว (Draney & Wilson, 2011; Wilson & Draney, 2002) ซึ่งเท่ากับว่าสามารถใช้วิธีนี้ร่วมกับวิธีกำหนดคะแนนจุดตัดอื่นได้

นอกจากนี้วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์สามารถใช้เป็นฐานให้กับวิธีกำหนดคะแนนจุดตัดอื่นๆ ที่เน้นการประมาณพารามิเตอร์ข้อสอบและผู้สอบ เนื่องจากวิธีดังกล่าวนำโมเดลการตอบสนองข้อสอบและแผนที่ภาวะสันนิษฐานมาใช้เป็นส่วนประกอบสำคัญของการกำหนดคะแนนจุดตัด โดยเฉพาะอย่างยิ่งกลุ่มโมเดลการวัดในตระกูลราสช์ เช่น พารามิเตอร์ข้อสอบและผู้สอบที่ประมาณได้จากโมเดลลอจิสติกเบื้องต้น (simple logistic model: Rasch, 1960) โมเดลมาตราประมาณค่า (rating scale model: Andrich, 1978) โมเดลให้คะแนนบางส่วน (partial credit model: Masters, 1982) โมเดลการแบ่งเชิงจัดอันดับ (ordered partition model: Wilson, 1992) โมเดลหลายฟาเซต (Many facets model: Fischer, 1983; Linacre, 1990) โมเดลขั้นพัฒนาการ (Saltus model; Wilson, 1989) และโมเดลชุดข้อสอบที่สัมพันธ์กัน (item bundles model: Wilson & Adams, 1995) ซึ่งสามารถนำเสนอในรูปแบบของแผนที่พารามิเตอร์ความยากข้อสอบและความสามารถแฝงผู้สอบที่เชื่อมต่อกับแผนที่ภาวะสันนิษฐานหรือคำบรรยายระดับความสามารถได้

แผนที่ภาวะสันนิษฐานและโมเดลการวัดของราสช์

วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์จะเชื่อมโยงสารสนเทศหลายส่วนเข้าด้วยกัน ได้แก่ ค่าพารามิเตอร์ความสามารถผู้สอบ ค่าพารามิเตอร์ความยากข้อสอบ การแจกแจงค่าพารามิเตอร์ข้อสอบและผู้สอบ คุณลักษณะข้อสอบ คุณลักษณะผู้สอบและความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบถูก สารสนเทศเหล่านี้ช่วยให้คะแนนจุดตัดขั้นต้นสามารถตีความได้ทั้งแบบอิงเกณฑ์ (criterion-referenced) และอิงกลุ่ม (norm-referenced) (Draney & Wilson, 2011; Wilson & Draney, 2002; Wyse, 2013) วิธีทำแผนที่ภาวะสันนิษฐานนำมโนทัศน์ของสเกลการตอบสนองข้อสอบ (item response scale) ที่ได้จากการสร้างโมเดลการวัดในตระกูลราสช์ (Rasch-family item response model) มาใช้เป็นแนวทางกำหนดสิ่งที่นักเรียนจะรู้และจะทำได้ ทำให้กระบวนการเลือกค่าที่จะเป็นคะแนนจุดตัดซึ่งอยู่บนสเกลตัวแปรแฝง (latent trait scale) มีความชัดเจนและตรงไปตรงมาบนพื้นฐานของผลวิเคราะห์ข้อมูล (Draney & Wilson, 2011; Wilson & Draney, 2002) แทนที่จะเป็นคะแนนจุดตัดที่จะเกิดขึ้นจากการเปรียบเทียบ

วัตถุประสงค์สำคัญของการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์คือค่าพารามิเตอร์ความยากข้อสอบ (δ) และค่าประมาณความสามารถแฝงผู้สอบ (θ) ที่ได้จากโมเดลการวัดของราสช์ (Rasch, 1960; Wright & Stone, 1979) โมเดลราสช์แบบดั้งเดิมนั้นจะสมมติให้

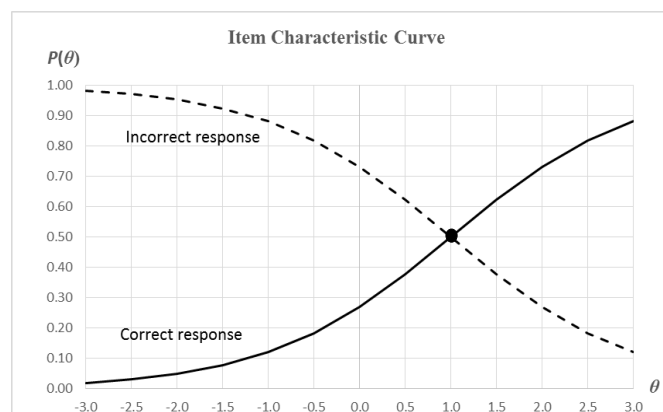
ข้อสอบทุกข้อในแบบสอบวัดสิ่งเดียวกันที่ประกอบด้วยมิติเดียว ดังนั้นโมเดลนี้จึงอาศัยหลักการความเป็นเอกมิติของแบบสอบหรือสิ่งที่มุ่งวัด มโนทัศน์เกี่ยวกับความเป็นเอกมิติในโมเดลราส์สามารถอธิบายให้เข้าใจง่ายขึ้นด้วยข้อสอบที่ให้คะแนนแบบสองค่า (dichotomous scoring) เช่นเดียวกับข้อสอบ ONET โดยคำตอบถูกมีค่าเท่ากับ 1 และคำตอบผิดเท่ากับ 0 ความน่าจะเป็นของผู้สอบ n ($n = 1, 2, \dots, N$) ที่ตอบข้อสอบ i ($i = 1, 2, \dots, I$) ถูกสามารถคำนวณได้จากสมการ (1)

$$P(X_i = 1|\theta_n, \delta_i) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \quad (1)$$

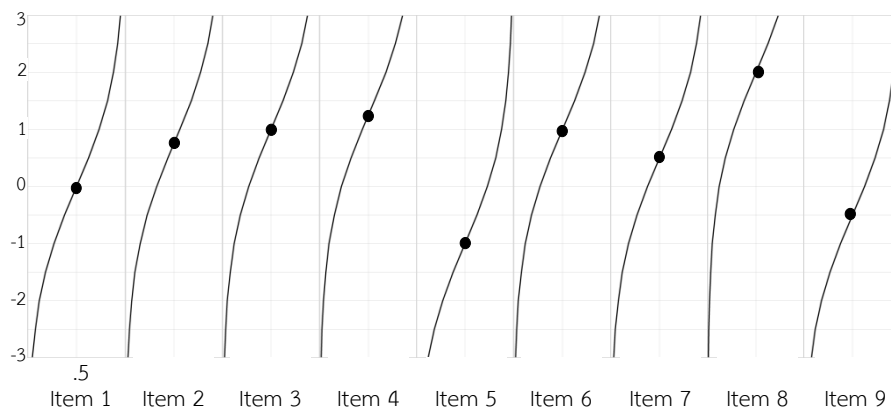
ในทางตรงกันข้าม ความน่าจะเป็นที่ผู้สอบ n ที่ตอบข้อสอบ i ผิดสามารถคำนวณได้จากสมการ (2)

$$P(X_i = 0|\theta_n, \delta_i) = 1 - P(X_i = 1|\theta_n, \delta_i) = \frac{1}{1 + \exp(\theta_n - \delta_i)} \quad (2)$$

และเมื่อนำค่าความน่าจะเป็นทั้งสองมาพอร์ตโค้งลักษณะข้อสอบ (item characteristic curve) ดังภาพที่ 2.6 โดยเส้นทึบเป็นความน่าจะเป็นของการตอบถูก ส่วนเส้นประเป็นความน่าจะเป็นของการตอบผิด จุดตัดของกราฟทั้งสองตรงกับความน่าจะเป็น .5 หมายความว่าผู้สอบที่มีค่าประมาณความสามารถเท่ากับ 1.0 มีโอกาส 50% ที่จะตอบข้อสอบ i ได้ถูกต้อง ซึ่งก็คือค่าพารามิเตอร์ความยากของข้อสอบ i เท่ากับ 1.0 ดังนั้นจะเห็นว่าค่าประมาณความสามารถและค่าพารามิเตอร์ความยากข้อสอบเป็นค่าที่อยู่สเกลลอจิทเดียวกัน สมมติว่าแบบสอบฉบับหนึ่งประกอบด้วยข้อสอบจำนวน 9 ข้อ เมื่อนำโค้งคุณลักษณะข้อสอบทั้งหมดมาเรียงต่อกันจะมีลักษณะดังภาพที่ 2.7 โดยมี • เป็นจุดบอกตำแหน่งของ δ_i บนสเกลความสามารถที่อยู่ทางซ้ายมือ สารสนเทศที่ได้จากภาพที่ 2.7 นี้แสดงให้เห็นถึงการกระจายของ δ_i



ภาพที่ 2.6 ตัวอย่างโค้งคุณลักษณะข้อสอบ



ภาพที่ 2.7 ฟังก์ชันการตอบสนองข้อสอบจำนวน 9 ข้อ

ความสัมพันธ์อย่างง่ายระหว่าง θ_n และ δ_i สามารถพิจารณาได้จากอัตราส่วนออดส์ (odds ratio, ω) ของผู้สอบ n ต่อการตอบข้อสอบ i ได้ถูกต้อง แสดงดังสมการ (3)

$$\omega_i = \frac{P(X_i = 1|\theta_n, \delta_i)}{P(X_i = 0|\theta_n, \delta_i)} = \frac{1 + \exp(\theta_n - \delta_i)}{1/1 + \exp(\theta_n - \delta_i)} = \exp(\theta_n - \delta_i) \quad (3)$$

และลอการิทึมของอัตราส่วนออดส์ในสมการ (3) แสดงดังสมการ (4)

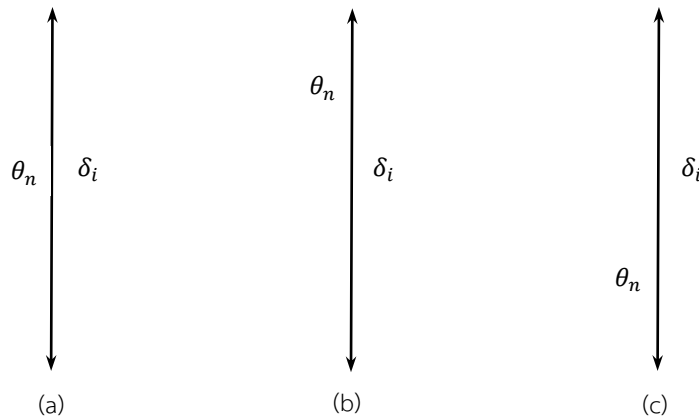
$$\log(\exp(\theta_n - \delta_i)) = \theta_n - \delta_i \quad (4)$$

เมื่อพิจารณาสมการ (4) พบว่ามีพารามิเตอร์ที่สนใจสองตัวคือ θ_n และ δ_i และน่าสังเกตว่าอัตราส่วนออดส์ของเหตุการณ์ใดๆ คืออัตราส่วนของโอกาสที่จะเกิดเหตุการณ์หนึ่งต่อโอกาสที่จะไม่เกิดเหตุการณ์นั้น ดังนั้นเมื่อหาลอการิทึมของอัตราส่วนออดส์ดังสมการ (4) จึงเป็นการเปรียบเทียบระหว่างความน่าจะเป็นของการตอบถูกกับการตอบไม่ถูกของข้อสอบแต่ละข้อ ลอการิทึมของอัตราส่วนออดส์เรียกว่า ลอจิท (logit) ดังนั้นสมการ (4) สามารถเขียนใหม่ดังสมการ (5)

$$\text{logit}(1:0) = \theta_n - \delta_i \quad (5)$$

สมการ (5) แสดงให้เห็นถึงความสัมพันธ์อย่างง่ายระหว่าง θ_n และ δ_i โดยค่าพารามิเตอร์ทั้งสองนี้มีหน่วยการวัดเดียวกันที่เรียกว่า ลอจิท และอยู่บนสเกลเดียวกันด้วย เมื่อกำหนดให้มีเส้นจำนวนใดๆ ที่แสดงสเกลลอจิทดังภาพที่ 2.8 ก็จะสามารถระบุตำแหน่งของ θ_n และ δ_i ได้ ถ้าตำแหน่งของ θ_n อยู่สูงกว่าและ δ_i แสดงว่าความสามารถของผู้สอบ n มากกว่าความยากของข้อสอบ i นั่นคือ ผู้สอบ n มีโอกาสมากกว่า 50% ที่จะตอบข้อสอบ i ได้ถูกต้องดังภาพที่ 2.8(b) ในทางตรงกันข้าม ถ้าตำแหน่งความสามารถผู้สอบ n อยู่ต่ำกว่าดังภาพที่ 2.8(c) หมายความว่า ผู้สอบ n มี

โอกาสน้อยกว่า 50% ที่จะตอบข้อสอบ i ได้ถูกต้อง ส่วนภาพที่ 2.8(a) หมายความว่าผู้สอบ n มีโอกาสตอบข้อสอบ i ได้ถูกเท่ากับ 50%



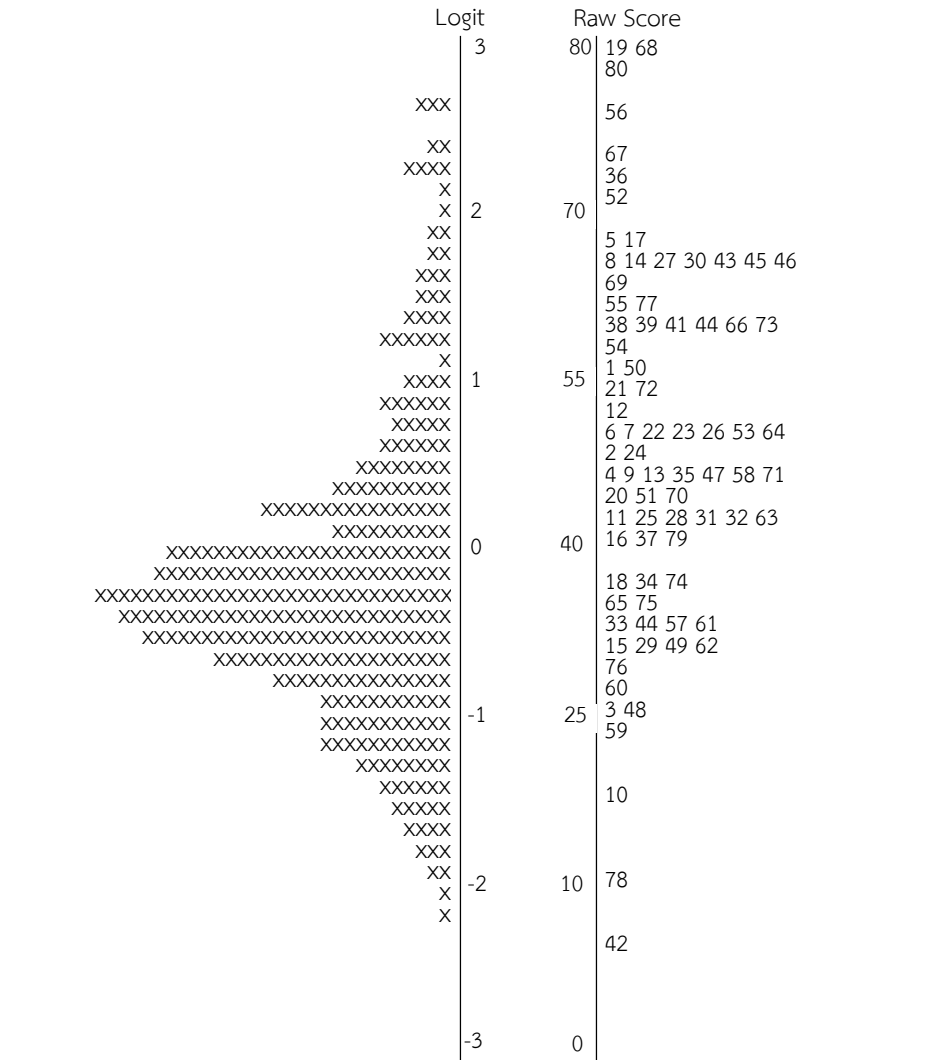
ที่มา: Constructing measures: An item response modeling approach (Wilson, 2005)

ภาพที่ 2.8 ความสัมพันธ์ระหว่างค่าประมาณความสามารถแฝงกับค่าพารามิเตอร์ความยากข้อสอบ

เมื่อพิจารณาสมการ (5) และภาพที่ 2.8 พบว่าความสัมพันธ์ระหว่างตำแหน่งค่าประมาณความสามารถแฝงและตำแหน่งค่าความยากข้อสอบเป็นแบบเส้นตรงและเมื่อนำมโนทัศน์นี้ไปใช้อธิบายความเป็นเอกมิติของแบบสอบจะพบว่า แบบสอบที่ประกอบด้วยข้อสอบทุกข้อวัดสิ่งเดียวกันที่มีเพียงมิติเดียว สเกลลอจิทที่แสดงความสัมพันธ์ระหว่าง θ_n และ δ_i จะมีเพียงเส้นเดียวด้วย ไม่ว่าแบบสอบนั้นจะประกอบด้วยข้อสอบจำนวนกี่ข้อก็ตาม ภาพที่ 2.9 แสดงการแจกแจงของค่าประมาณความสามารถแฝงผู้สอบ θ จำนวน N คนที่ทำแบบสอบฉบับหนึ่งซึ่งประกอบด้วยข้อสอบจำนวน I ข้อ กับค่าความยาก δ ของข้อสอบเหล่านั้นเรียกว่าแผนที่ข้อสอบและผู้สอบ (item-person map) หรือแผนที่ตัวแปร (variable map) (Bond & Fox, 2007) หรือแผนที่การตอบสนองข้อสอบ (item response map) หรือแผนที่ของ Wright (Wright map) ตามชื่อของ Benjamin D. Wright ซึ่งเป็นผู้ริเริ่มใช้แผนภาพดังกล่าว (Wilson, 2005; Wilson & Draney, 2002)

สมมติว่ามีนักเรียนจำนวน 300 คน ทำแบบสอบฉบับหนึ่งซึ่งประกอบด้วยข้อสอบจำนวน 80 ข้อ เนื่องจากคะแนนของข้อสอบแต่ละข้อเป็นแบบสองค่า คะแนนรวมต่ำสุดของแบบสอบฉบับนี้จึงเท่ากับ 0 และสูงสุดเท่ากับ 80 เมื่อประมาณค่า θ_n ($n = 1, 2, \dots, 300$) และ δ_i ($i = 1, 2, \dots, 80$) แผนที่ของ Wright จะมีลักษณะดังภาพที่ 2.9 สัญลักษณ์ x ที่อยู่ทางด้านซ้ายของแผนที่แสดงตำแหน่งของค่าประมาณความสามารถแฝง $\theta_{n(n=1,2,\dots,300)}$ บนสเกลลอจิทที่สัมพันธ์กับคะแนนรวม (total raw score) ส่วนหมายเลขที่อยู่ทางด้านขวาของแผนที่แทนลำดับที่ข้อสอบในแบบสอบและตำแหน่งของค่าความยาก $\delta_{i(i=1,2,\dots,80)}$ ข้อสอบแต่ละข้อบนสเกลลอจิทที่สัมพันธ์กับคะแนนรวม

ตัวอย่างเช่นค่าพารามิเตอร์ความยากข้อสอบข้อ 3, 48 และ 59 มีค่าประมาณ -1.0 ลอจิต ซึ่งตรงกับค่าประมาณความสามารถของนักเรียนที่ -1.0 ลอจิตหรือคะแนนดิบรวม 25 คะแนน นั่นคือนักเรียนที่มีค่าประมาณความสามารถน้อยกว่า -1.0 จำนวน 52 คนจะมีโอกาสตอบข้อสอบข้อ 3, 48 และ 59 ผิดมากกว่า 50% ในขณะที่นักเรียนที่เหลือมีโอกาสตอบข้อ 3, 48 และ 59 ถูกอย่างน้อย 50% ยิ่งนักเรียนมีค่าประมาณความสามารถมากกว่า -1.0 มากเท่าไรก็จะยิ่งมีโอกาสตอบถูกมากยิ่งขึ้นเท่านั้น



ภาพที่ 2.9 ตัวอย่างของแผนที่ของ Wright จากโมเดลเอกมิตี

ในบางสถานการณ์ที่โมเดลพหุมิติมีความเหมาะสมมากกว่าโมเดลเอกมิตี จำนวนสเกลลอจิตและจำนวนของแผนที่ของ Wright จะมีจำนวนเท่ากับจำนวนมิติย่อย ตำแหน่งค่าประมาณความสามารถแฝง θ ของนักเรียนคนเดียวกันบนสเกลลอจิตของแต่ละมิติย่อยไม่จำเป็นต้องเหมือนกัน ในทำนองเดียวกันตำแหน่งความยากข้อสอบ δ ข้อเดียวกันที่วัดมากกว่าหนึ่งมิติย่อยก็ไม่จำเป็นต้องเหมือนกันเช่นกัน ตัวอย่างเช่น จากภาพที่ 2.8 นักเรียนคนเดียวกันทำข้อสอบจำนวนสามข้อ ซึ่งแต่ละ

ข้อวัดคนละมิติย่อย ค่าประมาณความสามารถแฝง θ ของผู้สอบคนนี้ก็จะมีสามค่าและตำแหน่งของ θ ก็ไม่จำเป็นที่จะต้องเหมือนกัน แม้ว่าข้อสอบทั้งสามข้อจะมีค่าความยาก δ เท่ากันก็ตาม

โมเดลการวัดที่ใช้ในการประมาณค่าความสามารถแฝง θ และค่าความยากข้อสอบ δ เป็นกรอบโมเดลที่เรียกว่า โมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์เชิงสุ่มจากการวัดแบบหลายมิติ (Multidimensional Random Coefficients Multinomial Logit model) หรือโมเดล MRCML (Adam, Wilson, & Wang, 1997) โดยกรอบโมเดลหมายถึงโมเดลทั่วไปที่สามารถดัดแปลงหรือเปลี่ยนแปลงข้อกำหนดลักษณะโมเดล (model specification) ที่อยู่ในรูปเมทริกซ์ให้สอดคล้องกับวัตถุประสงค์ของการวิเคราะห์และธรรมชาติของข้อมูลแล้วทำให้ได้โมเดลใหม่ เช่น เมื่อระบุข้อกำหนดลักษณะโมเดลว่าสิ่งที่สนใจจะวัดนั้นประกอบด้วยหลายมิติที่สัมพันธ์กัน ให้คะแนนคำตอบหลายค่าและมีข้อสอบจำนวนหนึ่งที่ไม่ได้เป็นอิสระกัน เช่น ใช้โจทย์คำถามร่วมกันและใช้คำตอบของข้อก่อนหน้าเพื่อหาคำตอบในข้อถัดไป โมเดลที่เกิดขึ้นใหม่จะเรียกว่า Partial credit-item bundles multidimensional model เป็นต้น ดังนั้นโมเดลการวัดในตระกูลราสส์สามารถวิเคราะห์ได้ด้วยกรอบโมเดล MRCML (Wu, Adam, Wilson, & Haldane, 2007)

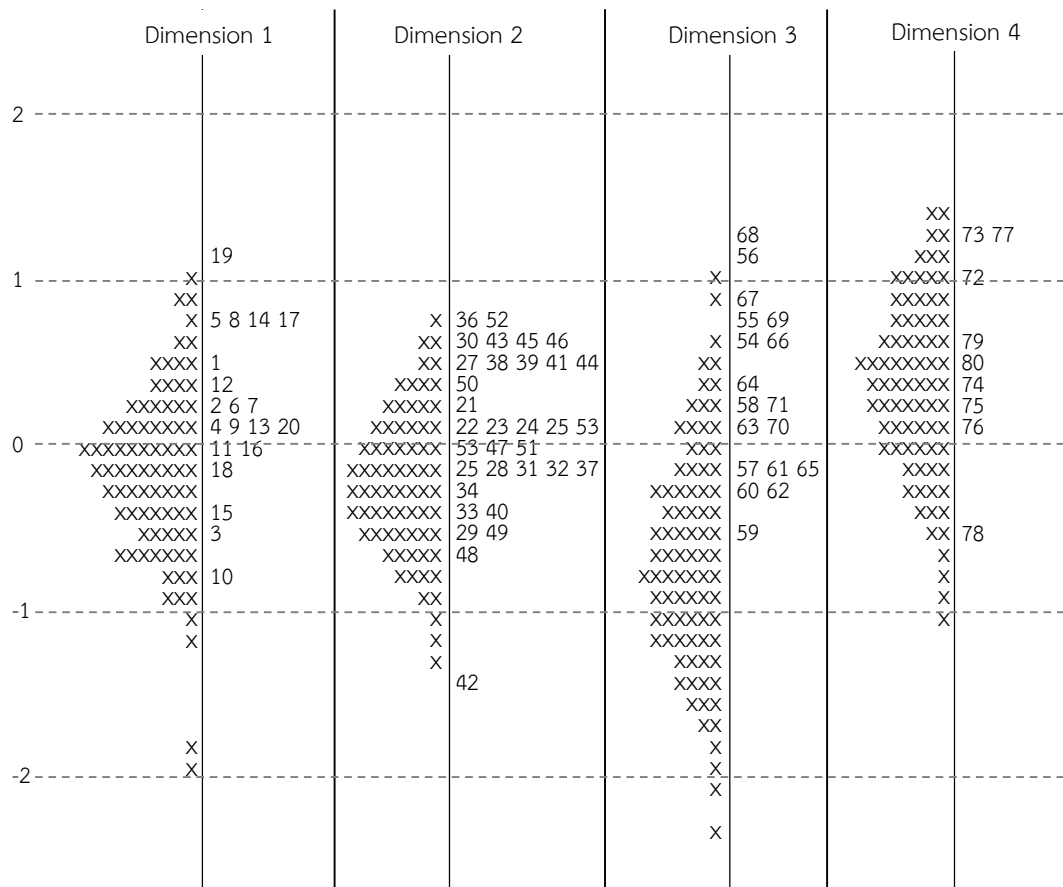
โมเดล MRCML พัฒนาจากโมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์เชิงสุ่มจากการวัดแบบมิติเดียว (Random Coefficients Multinomial Logit) หรือโมเดล RCML (Wilson & Adams, 1995; Wilson & Wang, 1995) ดังนั้นโมเดล MRCML สามารถวิเคราะห์ได้ทั้งโมเดลเอกมิติและโมเดลพหุมิติ ส่วนโมเดล RCML วิเคราะห์ได้เฉพาะโมเดลเอกมิติเท่านั้น โมเดล MRCML จะสมมติให้มิติย่อย d ($d = 1, 2, \dots, D$) ส่งผลต่อความสามารถแฝง θ ที่ใช้ข้อสอบ i ที่วัดมิติย่อย d ซึ่งก็คือคำตอบของผู้สอบ n ที่มีต่อข้อสอบ i ดังนั้นลอการิทึมของอัตราส่วนออดส์ของความน่าจะเป็นของคำตอบประเภท k ของข้อสอบ i (P_{ik}) เปรียบเทียบกับความน่าจะเป็นของคำตอบประเภท $k-1$ (P_{ik-1}) จะอยู่ในรูปฟังก์ชันเชิงเส้นตรงของความสามารถแฝงในแต่ละมิติย่อย d (θ_d) กับค่าความยากสัมพัทธ์ (relative difficulty) ของข้อสอบ i ที่มีคำตอบประเภท k (δ_{ik}) ดังสมการ (6)

$$\log \left(\frac{P_{nik}}{P_{nik-1}} \right) = \theta_{nd} - \delta_{ik} \quad (6)$$

θ_{nd} ในสมการ (6) คือความสามารถแฝงผู้สอบ n ในมิติย่อย d ซึ่งวัดได้ด้วยข้อสอบ i ดังนั้นผลของการระบุตำแหน่งค่าความยากข้อสอบ δ_i จะสัมพันธ์กับมิติย่อย d ค่า δ_{ik} ซึ่งโดยทั่วไปเรียกว่าค่าความยากลำดับขั้น (step difficulty หรือ threshold) เมื่อข้อสอบข้อนั้นให้คะแนนแบบหลายค่าหรือเรียกว่าค่าความยากข้อสอบเมื่อข้อสอบข้อนั้นให้คะแนนแบบสองค่า จะบ่งชี้ว่าข้อสอบ i นั้นยากหรือง่ายสำหรับผู้สอบ n โดยพิจารณาจากคำตอบประเภท k ของผู้สอบคนนั้นเปรียบเทียบกับคำตอบประเภท $k-1$

ค่าความสามารถแฝง θ_{nd} และค่าความยาก δ_i ที่ประมาณได้จากโมเดล MRCML สามารถนำมาสร้างเป็นแผนที่ของ Wright ได้เช่นเดียวกับการวิเคราะห์แบบเอกมิติ จากตัวอย่างข้างต้น สมมติ

ว่ามีนักเรียนจำนวน 300 คนทำข้อสอบจำนวน 80 ข้อ โดยแบ่งข้อสอบออกเป็นสี่มิตีย่อย ข้อสอบแต่ละข้อวัดหนึ่งมิติและให้คะแนนแบบสองค่า แผนที่ของ Wright จะมีลักษณะดังภาพที่ 2.10



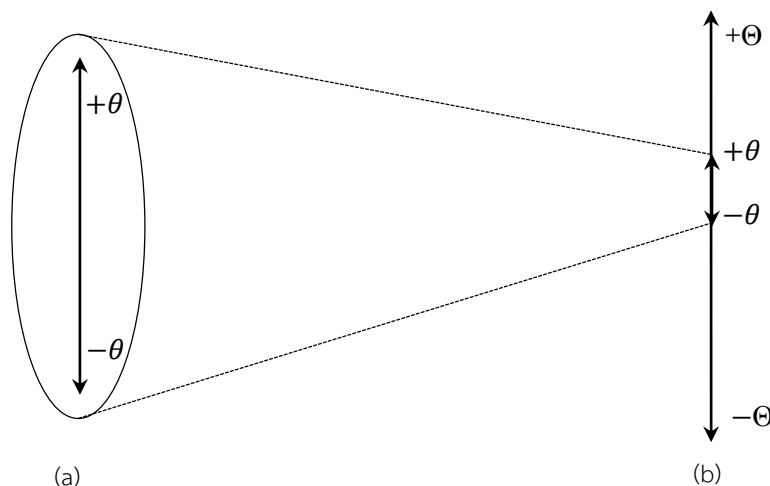
ภาพที่ 2.10 ตัวอย่างของแผนที่ของ Wright จากโมเดลพหุมิติ

แผนที่ของ Wright ดังภาพที่ 2.9 และ 2.10 ให้สารสนเทศแตกต่างกัน และเป็นแหล่งข้อมูลสำคัญสำหรับการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ จะเห็นว่าวิธีดังกล่าวนี้เชื่อมโยงหลักการของทฤษฎีการตอบสนองข้อสอบเข้ามาเป็นส่วนหนึ่งในกระบวนการกำหนดคะแนนจุดตัด และแม้ว่าจากตัวอย่างแผนที่ของ Wright จะไม่ได้แสดงให้เห็นถึงค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถแฝงและความยากข้อสอบ ซึ่งเป็นไปตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยาปี 2014 สารสนเทศนี้ก็สามารถนำมาพอร์ตลงในแผนที่ของ Wright ได้ด้วย และในกรณีที่แบบสอบประกอบด้วยรูปแบบข้อสอบหลายแบบ เช่น ข้อสอบแบบหลายตัวเลือก กับข้อสอบแบบเติมคำตอบ ข้อมูลดังกล่าวนี้ก็สามารถเพิ่มลงไปในพื้นที่ของ Wright ได้เช่นกัน ทำให้เห็นถึงผลของรูปแบบข้อสอบที่ส่งผลต่อความยากข้อสอบและส่งผลกระทบต่อค่าที่จะกำหนดเป็นคะแนนจุดตัดขั้นต้นด้วย

การสร้างคำบรรยายระดับความสามารถด้วยวิธีทำแผนที่ภาวะสันนิษฐาน

แม้ว่า Schulz (2013) กล่าวถึงข้อดีของการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ว่า ในสถานการณ์จริงที่ไม่มีคำบรรยายระดับความสามารถมาก่อน ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจสามารถกำหนดคะแนนจุดตัดได้โดยพิจารณาจากแผนที่ของ Wright ที่เชื่อมโยงกับสารสนเทศอื่น ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจสามารถเลือกหรือปฏิเสธ หรือแก้ไขคะแนนจุดตัดในภายหลังได้เพราะแผนที่ภาวะสันนิษฐานสามารถทำหน้าที่แทนคำบรรยายระดับความสามารถได้และยังนำไปใช้พัฒนาข้อกำหนดลักษณะแบบสอบและนำไปใช้ในกระบวนการสร้างข้อสอบและให้คะแนนคำตอบได้ด้วย แต่อย่างไรก็ตามภายใต้กรอบการสร้างโมเดลภาวะสันนิษฐานนั้น สมควรอย่างยิ่งที่จะต้องมีการทำแผนที่ภาวะสันนิษฐานก่อนที่จะเริ่มสร้างข้อสอบเพื่อให้วัตถุประสงค์ของการสร้างแบบทดสอบและวัตถุประสงค์ของการกำหนดคะแนนจุดตัดสอดคล้องกันตั้งแต่เริ่มต้น

โมทัศน์ของการทำแผนที่ภาวะสันนิษฐานคือ การกำหนดโครงสร้างภายในให้กับตัวแปรแฝงความสามารถนักเรียนตามทฤษฎีที่เกี่ยวข้องให้อยู่บนสเกลการตอบสนองข้อสอบเช่นเดียวกับสเกลพารามิเตอร์ความสามารถผู้สอบ θ ที่ได้จากโมเดลการตอบสนองข้อสอบดังเช่นแกนทางซ้ายมือในภาพที่ 2.9 โดยทั่วไปแล้วแผนที่ภาวะสันนิษฐานประกอบด้วยข้อมูลสามส่วนคือ ขอบเขตของตัวแปรที่สนใจ คำบรรยายความสามารถของผู้ตอบ และคำบรรยายคำตอบ (Wilson, 2005) กล่าวคือ เส้นตรงภาวะสันนิษฐานหรือสเกลการตอบสนองข้อสอบที่มีหัวลูกศรทั้งสองด้านดังแสดงในภาพที่ 2.11a แสดงถึงปริมาณความสามารถของนักเรียนจากระดับน้อย (หัวลูกศรด้านล่าง: $-\theta$) สู่ระดับมาก (หัวลูกศรด้านบน: $+\theta$) ช่วงห่างระหว่างหัวลูกศรทั้งสองเป็นสิ่งที่สามารถกำหนดขึ้นเองเพื่อใช้แสดงถึงขอบเขตของตัวแปรความสามารถที่ไม่สามารถสังเกตได้โดยตรง แม้ว่าเส้นตรงนี้จะมีความยาวแตกต่างกันได้ แต่ขอบเขตของความสามารถที่ต้องการวัดเป็นเพียงส่วนหนึ่งของเส้นความต่อเนื่องที่แท้จริงของความสามารถแฝงนั้น (วีรภัทร์ สุขศิริ, 2557) (ดูภาพที่ 2.11b)



ภาพที่ 2.11 เส้นสมมติความต่อเนื่องของคุณลักษณะของตัวแปรแฝง

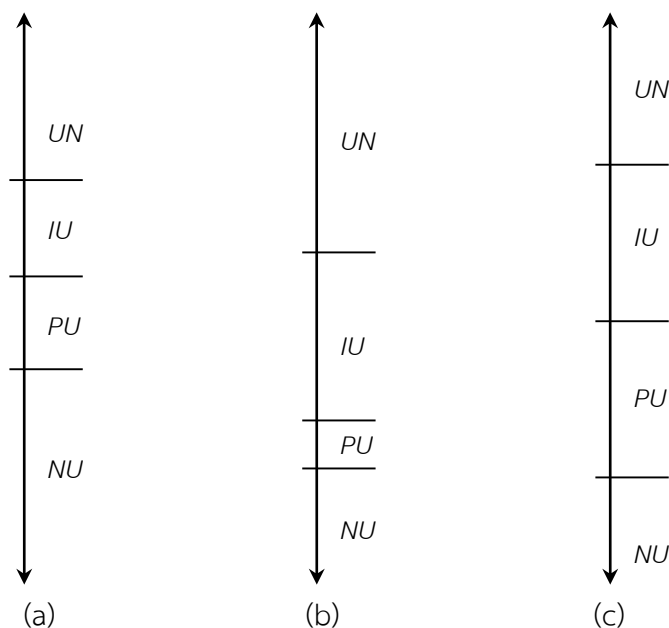
ระดับความเข้าใจ	คำบรรยายความสามารถของนักเรียน	คำบรรยายคำตอบของนักเรียน
เข้าใจ (UN): วิเคราะห์ ความสัมพันธ์ ระหว่างการคูณและการหารกับ หลักการทาง คณิตศาสตร์อื่นๆ	นักเรียนสามารถอธิบายสมบัติการคูณและการหารเข้าใจความสัมพันธ์ระหว่างการคูณและการหารในระบบสมการทางคณิตศาสตร์ สามารถสื่อสารทางคณิตศาสตร์ ใช้หลักการคูณและการหารแก้โจทย์ปัญหาที่ซับซ้อน เปรียบเทียบความแตกต่างของหลักการคูณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ เลือกใช้วิธีที่เหมาะสมในการแก้โจทย์ปัญหา	คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ถูกต้องในการเลือกใช้หลักการและวิธีการที่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อสนับสนุนหรือข้อโต้แย้งที่สมเหตุสมผลในการแก้โจทย์ปัญหา คำตอบที่เป็นไปได้บ่งชี้ว่า นักเรียนรู้คำตอบที่ถูกต้องและรู้ว่าคำตอบใดไม่ถูกต้อง
เข้าใจไม่สมบูรณ์ (UI): ใช้หลักการคูณและการหารแก้โจทย์ปัญหา	นักเรียนสามารถอธิบายสมบัติการคูณและการหารใช้ทักษะการสื่อสารทางคณิตศาสตร์และหลักการคูณและการหารแก้โจทย์ปัญหา เปรียบเทียบความแตกต่างของหลักการคูณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ สามารถจำแนกวิธีการหรือหลักการที่ไม่เหมาะสมในการแก้โจทย์ปัญหา	คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ถูกต้องในการเลือกใช้หลักการหรือวิธีการที่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อโต้แย้งหรือข้อสนับสนุนบางประการในการแก้โจทย์ปัญหา คำตอบที่เป็นไปได้บ่งชี้ว่า นักเรียนรู้คำตอบที่ถูกต้องและรู้เพียงบางคำตอบว่าเป็นคำตอบที่ไม่ถูกต้อง
เข้าใจบางส่วน (PU): อธิบายหลักการคูณและการหาร	นักเรียนสามารถอธิบายหลักการคูณและการหารได้เพียงบางส่วน แก้โจทย์ปัญหาได้อย่างจำกัดโดยอาจจะใช้หลักการบวกและการลบหรือหลักการทางคณิตศาสตร์อื่นๆ ใช้ทักษะการสื่อสารทางคณิตศาสตร์ได้จำกัด แต่อย่างไรก็ตาม นักเรียนสามารถจำแนกวิธีการหรือหลักการบางอย่างที่ไม่เหมาะสมในการแก้โจทย์ปัญหา	คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ไม่ถูกต้องจากการเลือกใช้หลักการที่ไม่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อสนับสนุนว่านักเรียนไม่สามารถใช้หลักการคูณและการหารแก้โจทย์ปัญหา มีข้อผิดพลาดในการสื่อสารทางคณิตศาสตร์ คำตอบที่เป็นไปได้บ่งชี้ว่านักเรียนไม่รู้คำตอบที่ถูกต้อง แต่รู้ว่าคำตอบใดไม่ถูกต้อง
ไม่เข้าใจ (NU): ตอบคำถามด้วย ข้อมูลที่เกี่ยวข้อง กับการคูณและ การหารเพียง เล็กน้อย	นักเรียนไม่สามารถใช้หลักการคูณและการหารแก้โจทย์ปัญหา ไม่สามารถตอบคำถามด้วยข้อมูลที่เกี่ยวข้องกับการคูณและการหาร นักเรียนอาจจะจดจำหลักการคูณและการหารได้ แต่ไม่สามารถนำมาใช้แก้โจทย์ปัญหา	คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อมูลที่เกี่ยวข้องกับการคูณและการหารเพียงเล็กน้อยหรือมีข้อมูลที่ไม่เกี่ยวข้องกับการแก้โจทย์ปัญหาจำนวนมาก คำตอบที่เป็นไปได้บ่งชี้ว่านักเรียนไม่รู้คำตอบที่ถูกต้องและไม่รู้ว่าคำตอบใดไม่ถูกต้อง นอกจากนี้ การไม่ตอบคำถาม เว้นว่าง และตอบคำถามไม่สมบูรณ์จนหาข้อสรุปไม่ได้ ถือว่าอยู่ในระดับนี้ด้วย

ที่มา: การสร้างแผนที่ตัวแปรเชิงทฤษฎี: แนวทางการสร้างโมเดลความคิดในระบบการประเมินที่ถูกต้องและน่าเชื่อถือ (วีรภัทร์ สุขศิริ, 2557)

ภาพที่ 2.12 ตัวอย่างคำบรรยายสรุประดับความเข้าใจเกี่ยวกับการคูณและการหาร

เส้นตรงในภาพที่ 2.11a นี้ทำหน้าที่เป็นแกนหลักของแผนที่ภาวะสันนิษฐานที่บอกปริมาณความสามารถที่ค่อยๆ เพิ่มขึ้นจากส่วนล่างไปถึงส่วนบนของเส้นตรง คุณลักษณะที่เกิดขึ้นระหว่างจุด $-θ$ ถึง $+θ$ สามารถแบ่งเป็นระดับต่างๆ ตามระดับความสามารถนักเรียนและรูปแบบการตอบสนองต่อข้อคำถาม จากนั้นจึงเขียนคำบรรยายความสามารถหรือคำบรรยายระดับความสามารถและคำบรรยายคำตอบที่สอดคล้องกับคำบรรยายความสามารถต่อไป (วีรภัทร์ สุขศิริ, 2557)

ภาพที่ 2.12 เป็นตัวอย่างแผนที่ภาวะสันนิษฐานความเข้าใจเรื่องการคูณและการหาร โดยมีคำบรรยายความสามารถและคำบรรยายคำตอบด้วย การเขียนคำบรรยายทั้งสองส่วนนี้เป็นการให้สารสนเทศในภาพรวมของการประเมินและมีความสำคัญที่สุดต่อการสร้างแผนที่ภาวะสันนิษฐานให้สมบูรณ์ รวมทั้งจะใช้สารสนเทศนี้เป็นแนวทางในการสร้างข้อคำถาม ใช้ตีความผลการวัดและสรุปอ้างอิงเกี่ยวกับความสามารถของนักเรียนต่อไป (วีรภัทร์ สุขศิริ, 2557) คำบรรยายคำตอบเป็นการให้ข้อมูลว่าคำตอบจะประกอบด้วยหลักฐานอะไรบ้างที่แสดงว่านักเรียนเข้าใจ เข้าใจไม่สมบูรณ์ เข้าใจบางส่วน และไม่เข้าใจตามแผนที่ภาวะสันนิษฐานที่สร้างขึ้น ในทำนองเดียวกันการสร้างคำบรรยายความสามารถของนักเรียนจึงเป็นการบรรยายว่านักเรียนมีคุณลักษณะอย่างไร รู้อะไรบ้างและทำอะไรได้ในแต่ละระดับ นอกจากนี้ ภาพที่ 2.4 ยังแสดงถึงตัวอย่างแผนที่ภาวะสันนิษฐานการให้เหตุผลทางวิทยาศาสตร์ซึ่งเมื่อนำแผนที่ด้านบนและด้านล่างมาจับคู่กันด้วยระดับความสามารถก็จะเห็นความสัมพันธ์ระหว่างโครงสร้างทางทฤษฎีของตัวแปรแฝงกับคำบรรยายคุณลักษณะของบุคคลที่อยู่ในแต่ละระดับและคำบรรยายคำตอบของบุคคลที่มีความสามารถอยู่ในระดับนั้นๆ



ที่มา: การสร้างแผนที่ตัวแปรเชิงทฤษฎี: แนวทางการสร้างโมเดลความคิดในระบบการประเมินที่ถูกต้องและน่าเชื่อถือ (วีรภัทร์ สุขศิริ, 2557)

ภาพที่ 2.13 ตัวอย่างตำแหน่งของเส้นแบ่งระดับความเข้าใจของนักเรียนเกี่ยวกับการคูณและการหาร

ภาพที่ 2.12 มีหลักการหนึ่งที่ซ่อนอยู่กล่าวคือ เส้นแบ่งความสามารถซึ่งมีมิติเดียวกันกับคะแนนจุดตัดระหว่างระดับความเข้าใจที่อยู่ติดกันนั้นไม่ได้ถูกกำหนดขึ้นเลย เนื่องจากผู้สร้างแบบสอบไม่สามารถระบุตำแหน่งของเส้นแบ่งแต่ละเส้นได้อย่างถูกต้องและแม่นยำจึงมีเพียงแนวคิดคร่าวๆ ว่า ความสามารถย่อนั้นน่าจะอยู่ตรงกับระดับความเข้าใจใดในแผนที่ภาวะสันนิษฐาน ดังนั้นจึงเป็นไปได้ที่แผนที่ภาวะสันนิษฐานที่ได้ภายหลังจากวิเคราะห์คำตอบข้อสอบอาจจะมีเส้นแบ่งความสามารถที่ทำให้ระดับเข้าใจและระดับไม่เข้าใจกว้างกว่าระดับอื่น (ภาพที่ 2.13a) หรืออาจจะมีเส้นแบ่งที่ทำให้ระดับเข้าใจบางส่วนและระดับไม่เข้าใจแคบ (ภาพที่ 2.13b) กว้างกว่าระดับอื่น หรืออาจจะเป็นไปได้ว่าแผนที่ภาวะสันนิษฐานมีเส้นแบ่งระดับความสามารถที่ทำให้ทุกระดับมีความกว้างใกล้เคียงกัน (ภาพที่ 2.13c) เป็นต้น ซึ่งสารสนเทศนี้จะช่วยให้การตัดสินใจเลือกค่าใดเป็นคะแนนจุดตัดมีความถูกต้องและชัดเจนมากขึ้น

จากที่อธิบายมาข้างต้นแผนที่ภาวะสันนิษฐานเป็นภาพของโครงสร้างภายในของตัวแปรแฝงซึ่งภาวะสันนิษฐานนี้เกิดขึ้นจากการค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้อง และอาจจะเกิดขึ้นจากข้อมูลเชิงคุณภาพที่ครูหรือนักวิจัยได้เก็บรวบรวมจากประสบการณ์ การสัมภาษณ์ การสังเกตหรือจากการแสดงความคิดเห็นของผู้เชี่ยวชาญ (Wilson, 2005) เป็นต้น แล้วนำข้อมูลดังกล่าวมาบรรยายคุณลักษณะของตัวแปรแฝงที่สนใจแทนที่จะแสดงว่าตัวแปรแฝงที่สนใจนั้นมีส่วนประกอบย่อยอะไรบ้าง แต่ไม่สามารถบรรยายเนื้อหาย่อยที่สะท้อนระดับหรือดีกรีของคุณลักษณะที่ทำหน้าที่เป็นสเกลที่อยู่ภายในแต่ละส่วนประกอบย่อยนั้นๆ การตรวจสอบคุณภาพเชิงตรรกะของแผนที่ภาวะสันนิษฐานที่สร้างขึ้นสามารถทำได้โดยตรงไปตรงมาและอยู่บนพื้นฐานของข้อมูลเชิงประจักษ์ เพราะภายหลังจากนำแผนที่ภาวะสันนิษฐานที่พัฒนาขึ้นจากทฤษฎีที่เกี่ยวข้องไปใช้สร้างข้อสอบและให้คะแนนคำตอบตามแผนที่ภาวะสันนิษฐานนั้นแล้ว เมื่อนำผลคะแนนคำตอบไปวิเคราะห์ก็จะได้ค่าพารามิเตอร์ความยากข้อสอบและเมื่อหาค่าเฉลี่ยพารามิเตอร์ความยากและพิจารณาการกระจายของพารามิเตอร์ดังกล่าว ก็จะเห็นว่าความกว้างของแต่ละระดับชั้นความสามารถที่ได้จากข้อมูลเชิงประจักษ์สอดคล้องกับโครงสร้างทางทฤษฎีหรือไม่ ตัวอย่างเช่น แผนที่ภาวะสันนิษฐานที่ระดับ NU มีความกว้างกว่าระดับอื่นๆ ซึ่งอาจจะไม่สอดคล้องกับโครงสร้างตั้งต้นที่ผู้สร้างแบบสอบกำหนดขึ้น ซึ่งความกว้างระดับชั้นนี้บ่งชี้ว่าความสามารถของนักเรียนในชั้นนั้นมีความแตกต่างกัน การแบ่งระดับความสามารถยังไม่ชัดเจนมากพอ และเป็นไปได้ว่าอาจจะมีระดับความสามารถอีกอย่างน้อยหนึ่งระดับแทรกระหว่างระดับ NU และระดับ PU เป็นต้น

ในทางตรงกันข้าม ภาพที่ 2.13b แสดงให้เห็นว่าความกว้างของระดับ PU แคบมากซึ่งหมายความว่าความสามารถของนักเรียนระดับ PU ไม่แตกต่างจากระดับ NU ดังนั้นอาจจะรวมระดับ PU และ NU เป็นระดับเดียวกันก็ได้ เมื่อมีการเปลี่ยนแปลงโครงสร้างของแผนที่ภาวะสันนิษฐานเป็นครั้งที่สอง ผู้สร้างแบบสอบก็จำเป็นต้องแก้ไขคำบรรยายความสามารถ คำบรรยายคำตอบ ข้อสอบและเกณฑ์การให้คะแนนคำตอบทุกครั้งให้สอดคล้องกับแผนที่ภาวะสันนิษฐานที่แก้ไขแล้ว (เนื่องจากสิ่งที่ต้องการวัดมีโครงสร้างเปลี่ยนไป และไม่เป็นไปตามทฤษฎีที่อ้างอิงไว้ในตอนแรก) และ

เมื่อวิเคราะห์ด้วยโมเดลการวัดของราซซ์อีกครั้ง ผลที่ได้จะเป็นหลักฐานเชิงประจักษ์อย่างดีและชัดเจนที่สุดเพื่อยืนยันว่าแผนที่ภาวะสันนิษฐานนั้นเหมาะสมที่จะใช้เป็นบรรยายความสามารถของนักเรียนหรือไม่ ผลวิเคราะห์ที่ขัดแย้งกับแผนที่ภาวะสันนิษฐานจะชี้แนะว่าผู้สร้างแบบสอบควรปรับปรุงแผนที่ภาวะสันนิษฐานนั้นอย่างไร ผลลัพธ์สุดท้ายที่ได้จากแผนที่ภาวะสันนิษฐานคือการจำแนกนักเรียนออกเป็นกลุ่มความสามารถที่ชัดเจน นักเรียนที่อยู่ในระดับชั้นความสามารถเดียวกันจะมีแนวโน้มสูงที่มีความสามารถเหมือนกัน และจะมีแนวโน้มต่ำที่จะมีความสามารถเหมือนกับนักเรียนในระดับชั้นอื่น

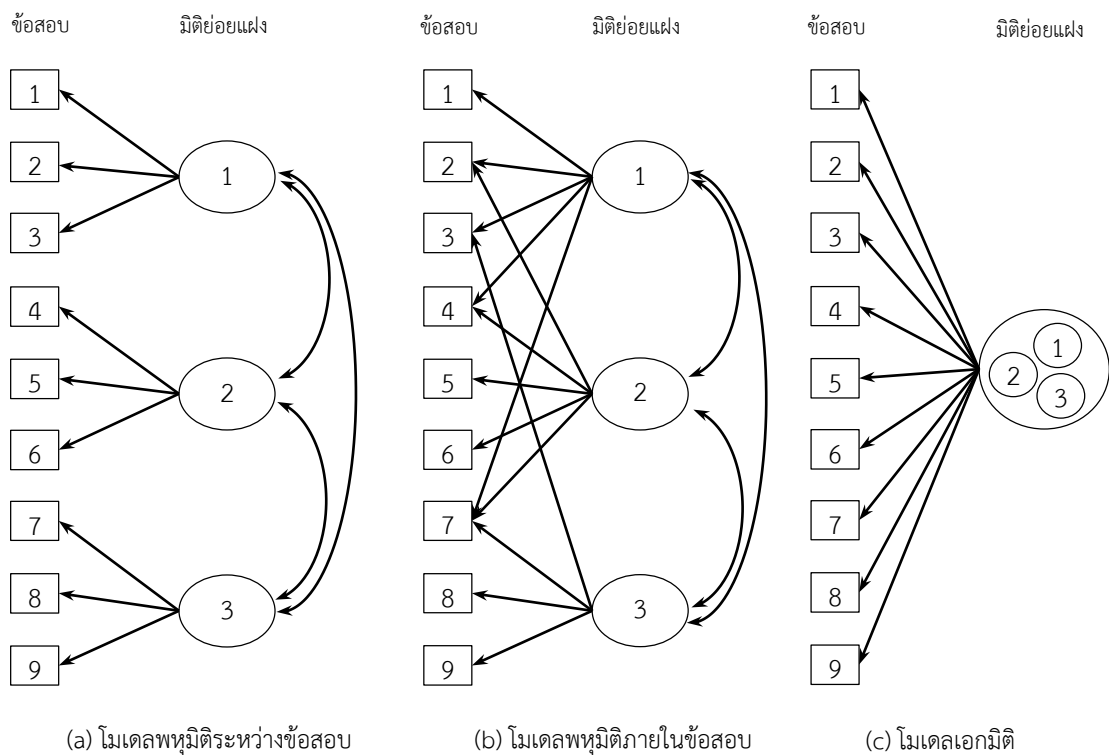
คุณลักษณะของโมเดลเอกมิติและโมเดลพหุมิติความสามารถของนักเรียน

ข้อตกลงเบื้องต้นประการหนึ่งของทฤษฎีการตอบสนองข้อสอบที่นิยมใช้อ้างอิงอย่างกว้างขวางคือความเป็นเอกมิติของข้อสอบ โดยข้อตกลงนี้สมมติให้ข้อสอบทั้งหมดในแบบสอบวัดคุณลักษณะแฝงที่สนใจเพียงมิติเดียวหรือวัดตัวแปรแฝงเพียงตัวเดียว (Hambleton, Swaminathan, & Roger, 1991; Rackase, 2009) โดยทั่วไปแล้ว การวัดคุณลักษณะแฝงก็คือการวัดความสามารถแฝงนักเรียนที่มีต่อข้อสอบ โดยความสามารถแฝงเกี่ยวกับคุณลักษณะใดๆ อาจจะประกอบด้วยมิตีย่อยที่สัมพันธ์กัน และมิตีย่อยเหล่านั้นสามารถพิจารณาเป็นมิติรวมเพียงหนึ่งเดียวที่ส่งผลต่อการตอบข้อสอบได้ ดังนั้นโมเดลการตอบสนองข้อสอบเอกมิติจึงเหมาะสมกับข้อตกลงเบื้องต้นนี้

อย่างไรก็ตาม ความเป็นเอกมิติของข้อสอบไม่สามารถใช้ได้ในทุกสถานการณ์ เมื่อตัวแปรแฝงทางจิตวิทยา เช่น ทักษะการแก้ปัญหา ความถนัดทางวิชาชีพ และผลสัมฤทธิ์ทางการเรียน มักจะประกอบด้วยมิตีย่อย อีกทั้งโครงสร้างแฝงภายในของตัวแปรเหล่านี้มีความซับซ้อนและแต่ละมิตีย่อยก็จะสัมพันธ์กันเองในระดับที่แตกต่างกันด้วย ทำให้นักเรียนจำเป็นต้องใช้ความสามารถมากกว่าหนึ่งอย่างเพื่อตอบข้อสอบที่วัดตัวแปรเหล่านี้ ดังนั้นข้อตกลงเบื้องต้นเกี่ยวกับความเป็นพหุมิติจึงถูกต้องและเหมาะสมมากกว่าความเป็นเอกมิติ ทั้งในด้านการวิเคราะห์และการตีความค่าประมาณความสามารถแฝงนักเรียนและค่าพารามิเตอร์ข้อสอบ โมเดลที่ใช้สำหรับประมาณค่าความสามารถแฝงจึงควรเป็นโมเดลการตอบสนองข้อสอบพหุมิติมากกว่าโมเดลเอกมิติ

โมเดลพหุมิติมีสองประเภทจำแนกตามแนวคิดในการสร้างข้อสอบและโครงสร้างของตัวแปร (Adams, Wilson, & Wang, 1997) โมเดลแรกเรียกว่า โมเดลพหุมิติระหว่างข้อสอบ (between-item multidimensionality model) โมเดลนี้จะประมาณค่าความสามารถแฝงผู้สอบด้วยข้อสอบในแบบสอบที่ประกอบด้วยมิตีย่อย โดยที่ข้อสอบหนึ่งข้อนั้นจะวัดคุณลักษณะย่อยเพียงหนึ่งมิติ เมื่อพิจารณาภาพที่ 2.14a พบว่า แบบสอบฉบับหนึ่งประกอบด้วยข้อสอบจำนวน 9 ข้อ โดยคุณลักษณะแฝงที่สนใจจะวัดนั้นประกอบด้วยสามมิตีย่อย ข้อสอบข้อที่ 1 ถึง 3 วัดมิตีย่อยที่ 1 ข้อสอบข้อที่ 4 ถึง 6 วัดมิตีย่อยที่ 2 และข้อสอบข้อที่ 7 ถึง 9 วัดมิตีย่อยที่ 3 โมเดลที่สองเรียกว่า โมเดลพหุมิติภายในข้อสอบ (within-item multidimensionality model) โมเดลนี้เหมาะสมกับแบบสอบที่มีข้อสอบ

บางข้อวัดมากกว่าหนึ่งมิตีย่อย เมื่อพิจารณาภาพที่ 2.14b จะพบว่าข้อสอบข้อที่ 1, 2, 3, 4 และ 7 วัดมิตีย่อยที่ 1 และในขณะเดียวกันข้อที่ 4 และ 7 ก็วัดมิตีย่อยที่ 2 และมิตีย่อยที่ 3 ตามลำดับด้วย ข้อสังเกตของโมเดลพหุมิติภายในข้อสอบคือ ไม่จำเป็นว่าข้อสอบทุกข้อต้องวัดอย่างน้อยสองมิตีย่อย ในแบบสอบฉบับนั้นอาจจะมีข้อสอบบางข้อที่วัดเพียงมิตีย่อยเดียว และมีบางข้อวัด 2 มิตีย่อย รวมทั้งเป็นไปได้ที่จะมีข้อสอบที่วัดครบทุกมิตีย่อยด้วยก็ได้ นอกจากนี้ข้อสอบทั้งหมดในแบบสอบฉบับเดียวกันก็ไม่จำเป็นต้องมีรูปแบบเหมือนกัน ตัวอย่างเช่น ข้อสอบที่วัดเพียง 1 หรือ 2 มิตีย่อยอาจจะเป็นข้อสอบแบบเลือกตอบ ในขณะที่ข้อสอบที่วัดหลายมิติอาจจะเป็นข้อสอบแบบเขียนตอบก็ได้



ภาพที่ 2.14 โมเดลการวัดแบบพหุมิติและโมเดลการวัดแบบเอกมิติ

สาระสำคัญของโมเดลพหุมิติที่แสดงในภาพที่ 2.14a และ 2.14b คือ เมื่อเปรียบเทียบกับโมเดลเอกมิติตั้งภาพที่ 2.14c ที่ข้อสอบแต่ละข้อไม่ได้ระบุจำเพาะว่าวัดมิตีย่อยใด พบว่า ข้อสอบที่สร้างขึ้นโดยอาศัยข้อตกลงเกี่ยวกับความเป็นเอกมิตินั้น ผู้สร้างแบบสอบจะพยายามสร้างข้อสอบทุกข้อให้วัดเพียงมิตีรวมแทนที่จะวัดมิตีย่อย เท่ากับว่า ผู้สร้างแบบสอบกำลังละเลยสารสนเทศเชิงวินิจฉัยความสามารถผู้สอบที่ควรจะได้รับจากข้อสอบ ข้อเท็จจริงของข้อสอบที่สร้างตามโมเดลเอกมิติ คือผู้สร้างแบบสอบจะสมมติว่าข้อสอบทั้งหมดในแบบสอบวัดเรื่องเดียวกันเพียงเรื่องเดียว เมื่อข้อสอบบางข้อมีแนวโน้มที่จะวัดมิตีย่อยอื่นด้วย ผู้สร้างแบบสอบก็จะพิจารณาว่าข้อสอบเหล่านั้นวัดเรื่องใดมากที่สุดเพียงอย่างเดียว (Hambleton, Swaminathan, & Roger, 1991) ข้อเท็จจริง

ดังกล่าวฝ่าฝืนธรรมชาติของข้อสอบและไม่สอดคล้องกับโครงสร้างภายในที่ซับซ้อนของคุณลักษณะแฝง รวมทั้งไม่สอดคล้องกับวัตถุประสงค์ของการทดสอบ และอาจทำให้โมเดลเชิงทฤษฎี (theoretical measurement model) ไม่กลมกลืนกับคำตอบข้อสอบด้วย ผลลัพธ์ที่ได้จากการวัดในลักษณะนี้จะขัดแย้งกับมาตรฐานการทดสอบทางการศึกษาและจิตวิทยาปี 2014 (AERA, APA, & NCME, 2014) ทำให้มีความตรงเชิงโครงสร้างต่ำ ความตรงเชิงตีความและการสรุปอ้างอิงเกี่ยวกับคุณลักษณะผู้สอบไม่ถูกต้องและผลการวัดไม่สามารถนำไปใช้ได้ตรงตามวัตถุประสงค์ของการสร้างแบบสอบด้วย (Borsboom, 2006; 2008; Kane, 2006; Markus & Borsboom, 2013; Messick, 1989; Michell, 2008; NRC, 2001; Wilson, 2005; Wilson et al., 2012)

หลักการเลือกโมเดลพหุมิติให้ถูกต้องคือ พิจารณาจากจุดประสงค์การสร้างแบบสอบที่ต้องการให้ข้อสอบมีคุณลักษณะอย่างไร ถ้าผู้สร้างข้อสอบมีแนวคิดทางทฤษฎีชัดเจนว่า สิ่งที่น่าสนใจวัดนั้นประกอบด้วยมิตีย่อยและแต่ละมิตีย่อยสะท้อนคุณลักษณะภายในที่แตกต่างกัน และผู้สร้างแบบสอบต้องการให้ข้อสอบบางข้อหรือทุกข้อวัดได้มากกว่าหนึ่งมิตีย่อย โมเดลที่เหมาะสมสำหรับประมาณค่าความสามารถแฝงผู้สอบควรจะเป็นโมเดลพหุมิติภายในข้อสอบ ผลวิเคราะห์ที่ได้จากโมเดลนี้ จะช่วยยืนยันความสอดคล้องกันระหว่างแนวคิดการสร้างข้อสอบในตอนต้นกับโมเดลที่ใช้วิเคราะห์คำตอบข้อสอบ ในบางครั้งที่ผลวิเคราะห์ไม่สามารถนำมาใช้ยืนยันความสอดคล้องดังกล่าวได้ แสดงว่าแนวคิดในการสร้างข้อสอบแตกต่างอย่างมีนัยสำคัญกับโมเดลที่ใช้ประมาณค่าความสามารถแฝง ผู้สร้างแบบสอบจึงไม่ควรแก้ปัญหาด้วยการเลือกใช้โมเดลพหุมิติระหว่างข้อสอบแทนที่จะเป็นโมเดลพหุมิติภายในข้อสอบตามที่ได้ออกแบบไว้ในตอนแรก วิธีการที่เหมาะสมคือผู้สร้างแบบสอบควรทบทวนอีกครั้งว่า ข้อสอบข้อใดไม่สามารถวัดบางมิตีย่อยตามที่ตั้งใจไว้ ข้อค้นพบลักษณะนี้จะประโยชน์อย่างมากต่อการปรับปรุงกระบวนการสร้างและพัฒนาข้อสอบ ส่งเสริมให้ผู้สร้างข้อสอบมีความเชี่ยวชาญในการออกแบบการวัดและมีความมั่นใจมากขึ้นในการระบุว่าข้อสอบแต่ละข้อนั้นวัดมิติตใดและไม่วัดมิติตใด

รูปแบบคำตอบ (response pattern) ที่เกิดขึ้นจะเป็นวัตถุบิตสำคัญสำหรับโมเดลพหุมิติเพื่อใช้แสดงความสัมพันธ์ระหว่างความน่าจะเป็นของคำตอบแต่ละรูปแบบกับตำแหน่งความสามารถแฝง (θ) ของผู้สอบแต่ละคนบนพื้นที่ระหว่างสเกลความสามารถย่อย (multidimensional space: Reckase, 2009) ตำแหน่งความสามารถแฝงผู้สอบ n สามารถนำเสนอในรูปเวกเตอร์ความสามารถแฝง (θ_n) = ($\theta_{1n}, \theta_{2n}, \dots, \theta_{dn}, \dots, \theta_{Dn}$) โดยที่ D ($d = 1, 2, \dots, D$) คือจำนวนมิตีย่อยแฝงของคุณลักษณะที่สนใจที่ใช้บรรยายความสามารถแฝงผู้สอบ n ที่ตอบข้อสอบของแต่ละมิตีย่อยนั้นๆ ค่าสมาชิก (element) ของเวกเตอร์จะบอกพิกัดหรือตำแหน่งความสามารถแฝงผู้สอบ n บนพื้นที่ระหว่างสเกลความสามารถย่อยจำนวน D มิติ ดังนั้นคำตอบของผู้สอบ n ที่มีต่อข้อสอบใดๆ โดยที่ข้อสอบนั้นวัดมิตีย่อยอย่างน้อยหนึ่งมิติจะเป็นตัวกำหนดพิกัดความสามารถแฝงผู้สอบคนนั้น และพิกัดดังกล่าวจะแตกต่างจากผู้สอบคนอื่นที่มีรูปแบบคำตอบของแต่ละมิตีย่อยไม่เหมือนกันด้วย

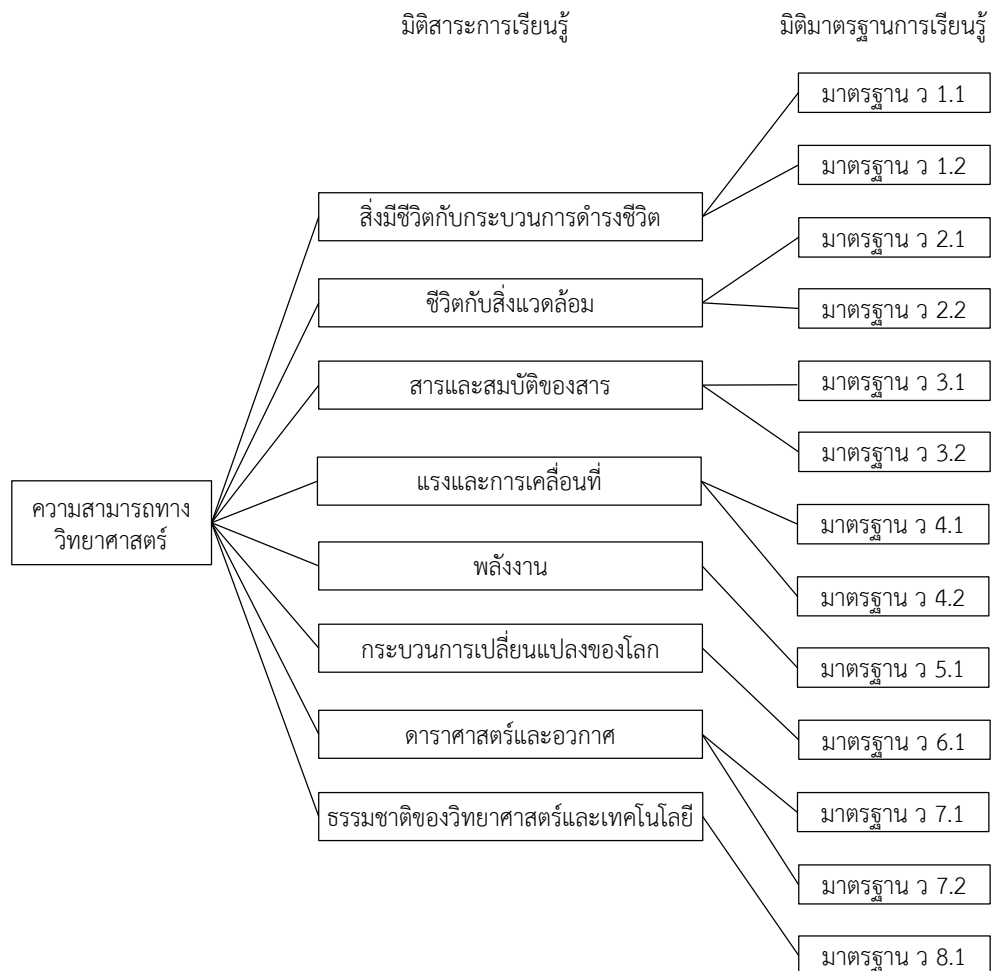
คุณสมบัติเด่นของโมเดลพหุมิติคือสามารถใช้ประมาณค่าความสามารถแฝงผู้สอบที่เกิดขึ้นจากการเรียนรู้ตามหลักสูตรหรือโปรแกรมการเรียนรู้ที่ประกอบด้วยเนื้อหาและ/หรือทักษะมากกว่าหนึ่งอย่าง นอกจากนี้โมเดลพหุมิตียังสามารถใช้กับการประเมินขนาดใหญ่ (large scale assessment) และการประเมินในชั้นเรียนได้เป็นอย่างดี เนื่องจากธรรมชาติของการประเมินทั้งสองประเภทจะประกอบด้วยเนื้อหาและคุณลักษณะหลายส่วนที่เกี่ยวข้องกัน ตัวอย่างเช่น การทดสอบความพร้อมทางการเรียนในระดับมหาวิทยาลัยที่ผู้สอบจะทำแบบสอบที่ประกอบด้วยข้อสอบวิชาวิทยาศาสตร์ คณิตศาสตร์ ภาษาอังกฤษและการให้เหตุผล เป็นต้น (Briggs & Wilson, 2003)

สำหรับการประเมินความสามารถทางวิทยาศาสตร์ของประเทศไทยในปัจจุบัน แบบสอบส่วนใหญ่จะอ้างอิงโครงสร้างความรู้และทักษะที่จำเป็นตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (กระทรวงศึกษาธิการ, 2551) โครงสร้างหลักสูตรดังกล่าวประกอบด้วยเนื้อหาและทักษะกระบวนการทางวิทยาศาสตร์มากกว่าหนึ่งมิติทำให้ลักษณะตามธรรมชาติของความสามารถทางการเรียนวิทยาศาสตร์เป็นพหุมิติด้วย ดังนั้นคุณลักษณะแฝงดังกล่าวก็สามารถแยกออกเป็นมิตีย่อยตามโครงสร้างของหลักสูตรแกนกลางได้เช่นกัน ภาพที่ 2.15 แสดงให้เห็นว่า ความสามารถทางวิทยาศาสตร์สามารถแบ่งออกเป็นรายการการเรียนรู้และรายมาตรฐานการเรียนรู้ได้ด้วย

เมื่อนำโครงสร้างความสามารถของนักเรียนดังภาพที่ 2.15 เชื่อมโยงเข้ากับข้อกำหนดลักษณะแบบสอบ ONET ที่สามารถเขียนให้อยู่ในรูปของโมเดล MRCML และข้อสอบ ONET แต่ละข้อถูกระบุไว้แล้วว่าวัดมาตรฐานการเรียนรู้ใดจากผู้สร้างแบบสอบ ONET ตั้งแต่ต้นจะพบว่า ไม่ว่าข้อสอบจะให้คะแนนแบบสองค่าหรือหลายค่า หรือมีข้อสอบบางข้อที่รวมเป็นกลุ่ม หรือมีข้อสอบบางข้อวัดเนื้อหามากกว่า 1 อย่าง หรือมีข้อสอบบางข้อวัดเนื้อหาเพียงหนึ่งอย่างแต่เนื้อหาที่วัดนั้นสัมพันธ์กับเนื้อหาส่วนอื่นๆ ในแบบสอบฉบับเดียวกัน หรือมีคุณลักษณะดังที่กล่าวมาสมอยู่ในแบบสอบฉบับเดียวกันก็ตาม สามารถใช้โมเดล MRCML ประมาณค่าพารามิเตอร์ผู้สอบและข้อสอบได้ทั้งหมด นอกจากนี้เมื่อพิจารณารายละเอียดของข้อมูลจำเพาะของแบบสอบ ONET พบว่า ผู้สร้างแบบสอบต้องการให้แบบสอบ ONET มีความเป็นพหุมิติมากกว่าเอกมิติ เช่น แบบสอบ ONET วิชาวิทยาศาสตร์ที่สถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) ต้องการให้แบบสอบดังกล่าววัดครอบคลุมทั้งมาตรฐานความรู้และมาตรฐานความสามารถ จากผลวิเคราะห์แบบสอบ ONET ที่ผ่านมา พบว่าการประมาณค่าความสามารถแฝงของนักเรียนยังให้ความสำคัญกับความเป็นเอกมิติของแบบสอบซึ่งขัดแย้งกับกระบวนการสร้างและพัฒนาแบบสอบของ สทศ. ค่าประมาณความสามารถแฝงจึงเป็นแบบภาพรวม ทำให้ไม่สามารถบ่งชี้ความสามารถของนักเรียนเป็นรายการการเรียนรู้หรือรายมาตรฐานการเรียนรู้ได้อย่างถูกต้องและเหมาะสม และถ้าหากนำสารสนเทศเหล่านี้ไปใช้กำหนดเป็นคะแนนจุดตัดก็ย่อมส่งผลเสียต่อความถูกต้องและความแม่นยำเช่นเดียวกัน

การตรวจสอบโครงสร้างความสามารถทางวิทยาศาสตร์ดังภาพที่ 2.15 กับโมเดล MRCML ที่สร้างข้อมูลจำเพาะของแบบสอบ ONET ว่ามีความเชื่อมโยงกันอย่างไร สามารถทำได้สองขั้นตอน ขั้นตอนแรกเป็นการวิเคราะห์ข้อมูลเชิงคุณภาพที่จะกำหนดให้ความสามารถนักเรียนแต่ละมาตรฐาน

การเรียนรู้เป็นคุณลักษณะแฝงที่สามารถวัดได้ด้วยข้อสอบ ONET ตามที่ผู้สร้างแบบสอบตั้งใจสร้าง เพื่อวัดมาตรฐานการเรียนรู้ต่างๆ ดังนั้นโมเดล MRCML ของแบบสอบ ONET วิชาวิทยาศาสตร์จึงประกอบด้วยมิติมาตรฐานย่อยจำนวน 13 มิติ และในทำนองเดียวกันก็สามารถสร้างโมเดล MRCML ที่สนใจมิติสาระการเรียนรู้ย่อยที่เกิดขึ้นจากการยุบรวมมาตรฐานการเรียนรู้ที่เกี่ยวข้องกันเข้าด้วยกัน ได้เป็นความสามารถแฝงสาระการเรียนรู้จำนวน 8 มิติ การตรวจสอบความเชื่อมโยงกันในขั้นตอนแรกนี้จึงเน้นไปที่การตัดสินใจของผู้เชี่ยวชาญเนื้อหาทางวิทยาศาสตร์และผู้สร้างแบบสอบเป็นสำคัญ ขั้นตอนที่สองเป็นการตรวจสอบเชิงปริมาณคือ เป็นการตรวจสอบความสอดคล้องกลมกลืนกันของโมเดลกับข้อมูลคำตอบข้อสอบ ผลวิเคราะห์ความสอดคล้องของโมเดล MRCML แต่ละโมเดลที่พัฒนาขึ้นในขั้นตอนแรกจะถูกนำมาเปรียบเทียบซึ่งกันและกันเพื่อหาโมเดลที่สอดคล้องกลมกลืนมากกว่า



ภาพที่ 2.15 โครงสร้างความสามารถทางวิทยาศาสตร์ตามหลักสูตรแกนกลางฯ พ.ศ. 2551

ตัวอย่างงานวิจัยที่เกี่ยวข้อง

ตัวอย่างงานวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ประกอบด้วยงานวิจัยของ Draney และ Wilson (2011) Lehrer, Kim, Ayers และ Wilson (2014) Wilson และ Draney (2002) และ Wyse, Bunch, Deville และ Viger (2013) โดยงานวิจัยเหล่านี้ไม่มีวัตถุประสงค์เพื่อเปรียบเทียบว่าวิธีดังกล่าวเป็นวิธีที่ดีกว่าวิธีอื่น ข้อค้นพบจึงไม่อ้างว่าวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์เป็นเพียงวิธีเดียวที่เชื่อมโยงกระบวนการกำหนดคะแนนจุดตัดเข้ากับกระบวนการพัฒนาแบบสอบ เนื่องจากยังมีวิธีของ Mislavy, Steinberg และ Almond (2003) วิธีของ Wang (2003) และวิธีของ Reckase (1998) ที่สามารถทำได้ แต่วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์เป็นเพียงวิธีเดียวที่ใช้แผนที่ภาวะสันนิษฐานเป็นแกนกลางสำคัญเพื่อเชื่อมโยงกระบวนการทั้งสอง และสามารถใช่วิธีดังกล่าวร่วมกับวิธีอื่นเพื่อเชื่อมโยงสารสนเทศต่างๆ มาใช้ประกอบการตัดสินใจที่จะเป็นคะแนนจุดตัด แนวทางการนำแผนที่ภาวะสันนิษฐานไปใช้กำหนดคะแนนจุดตัดจากตัวอย่างงานวิจัยข้างต้นมีมุมมองต่อความสามารถนักเรียนแบบตัวแปรพัฒนาการที่เริ่มต้นจากระดับพื้นฐานสู่ระดับสูงและแต่ละระดับจะมีรายละเอียดพฤติกรรมที่สัมพันธ์กับโครงสร้างพัฒนาการนั้น (Raymond, 2016) และพฤติกรรมเหล่านั้นจะแสดงออกมาจากการตอบสนองข้อสอบสารสนเทศเกี่ยวกับข้อสอบจึงสำคัญอย่างมากต่อการนิยามระดับความสามารถ

งานวิจัยของ Draney และ Wilson (2011) และ Wilson และ Draney (2002) เสนอว่ากระบวนการกำหนดคะแนนจุดตัดที่ไม่เชื่อมโยงกับกระบวนการทดสอบมีข้อจำกัดหลายประการและไม่เป็นไปตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยาของประเทศสหรัฐอเมริกา Wilson และ Draney (2002) แสดงให้เห็นว่าสามารถทำให้กระบวนการกำหนดคะแนนจุดตัดเป็นส่วนหนึ่งของระบบการทดสอบโดยใช้โมเดล MRCML ประมาณค่าพารามิเตอร์ความสามารถนักเรียนชั้นมัธยมศึกษาตอนปลาย วิชาสังคมศึกษา จากนั้นจึงนำเสนอค่าประมาณดังกล่าวในรูปของแผนที่ภาวะสันนิษฐานเชิงประจักษ์เพื่อแสดงว่าแบบสอบที่มีข้อสอบหลายรูปแบบส่งผลให้นักเรียนที่มีคะแนนสอบแตกต่างกันและมีโอกาสตอบถูกไม่เท่ากัน ทำให้ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจพิจารณาสารสนเทศดังกล่าวในระหว่างที่กำหนดคะแนนจุดตัด นอกจากนี้งานวิจัย Draney และ Wilson (2011) ได้ตรวจสอบซ้ำโดยใช้กลุ่มตัวอย่างขนาดใหญ่และใช้ข้อมูลคำตอบวิชาเคมีที่ซับซ้อนมากขึ้นทั้งในเรื่องของรูปแบบข้อสอบ ธรรมชาติของเนื้อหาวิชา และสัดส่วนข้อสอบในแต่ละตอน ข้อค้นพบที่ได้ช่วยยืนยันว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถเชื่อมโยงสารสนเทศหลายอย่างที่เกี่ยวข้องกับข้อสอบและนักเรียนเข้าด้วยกัน ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจสามารถนำสารสนเทศเหล่านั้นไปใช้ประกอบการตัดสินใจ รวมทั้งยังมีหลักฐานมาสนับสนุนแนวคิดของผู้มีอำนาจตัดสินใจด้วย

โครงการวิจัยเพื่อพัฒนาระบบการประเมินความสามารถด้านสถิตินักเรียนระดับประถมศึกษาของ Lehrer et al. (2014) แสดงให้เห็นว่าแผนที่ภาวะสันนิษฐานความสามารถทางสถิติสามารถออกแบบเป็นหลายมิติได้ รวมทั้งออกแบบเป็นโครงสร้างพัฒนาการจากระดับชั้นประถมศึกษาตอนต้น

ถึงตอนปลาย ระดับการเรียนรู้ของนักเรียนแต่ละชั้นปีจึงต่อเนื่องกันเป็นโครงสร้างของหลักสูตร ซึ่งเมื่อออกแบบข้อสอบให้สอดคล้องกับโครงสร้างดังกล่าวจะทำให้คะแนนสอบของนักเรียนสัมพันธ์กับโครงสร้างพัฒนาการนั้นด้วย และเมื่อนำคะแนนสอบไปใช้กำหนดเป็นคะแนนจุดตัด ผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจนอกจากจะเห็นความเชื่อมโยงกันของข้อมูลข้อสอบและข้อมูลนักเรียนแต่ละชั้นปีแล้วยังเห็นความเชื่อมโยงกันของความสามารถระหว่างชั้นปี และแนวโน้มความสามารถนักเรียนในแต่ละด้านด้วย คะแนนจุดตัดที่กำหนดเป็นมาตรฐานความสามารถนักเรียนจึงเหมาะสมกับนักเรียนแต่ละชั้นและลดความคลาดเคลื่อนในการตีความความสามารถของนักเรียนด้วย

งานวิจัยของ Wyse, Bunch, Deville และ Viger (2013) เสนอแนวทางการใช้วิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ร่วมกับวิธี Body of Work (Kingston, Kahl, Sweeney, & Bay, 2001; Kingston & Tiemann, 2012) เพื่อกำหนดคะแนนจุดตัด ข้อจำกัดโดยธรรมชาติของวิธี Body of Work คือ วิธีดังกล่าวเหมาะสมกับข้อสอบแบบเขียนตอบและชิ้นงานที่นักเรียนสามารถตอบได้อย่างอิสระ ครูไม่สามารถให้คะแนนชิ้นงานอย่างคงเส้นคงวา คะแนนที่ไม่คงเส้นคงวาและอาจจะไม่ตรงกับความสามารถจริงของนักเรียน ทำให้นักเรียนที่ได้คะแนนเท่ากันอยู่ต่างระดับความสามารถ นักเรียนที่มีความสามารถสูงอาจจะอยู่ในระดับความสามารถต่ำกว่าที่ควรจะเป็น ในขณะที่นักเรียนที่มีความสามารถปานกลางอาจอยู่ในระดับความสามารถที่สูงเกินจริงได้ นอกจากนี้ความแตกต่างของชิ้นงานยังทำให้เกิดช่องว่างของคะแนน เพราะครูไม่สามารถระบุให้ชัดเจนว่าชิ้นงานเหล่านั้นควรจะมีลำดับคะแนนต่อเนื่องกันอย่างไร ทำให้คะแนนบางช่วงไม่สามารถจับคู่กับชิ้นงานได้ รวมถึงปัญหาการตีความข้อเสนอแนะจากผู้เชี่ยวชาญเป็นคะแนนจุดตัดที่ผู้เชี่ยวชาญคนอื่นไม่เข้าใจหลักการคิดของผู้เชี่ยวชาญคนนั้นๆ อย่างไรก็ตามเมื่อนำวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์มาเชื่อมโยงข้อมูลข้อสอบไม่ว่าจะเป็นรูปแบบข้อสอบและการแจกแจงคะแนนเข้ากับหลักคิดและข้อเสนอแนะจากผู้เชี่ยวชาญแต่ละคน แล้วจึงตัดสินคะแนนจุดตัดพบว่า คะแนนจุดตัดสอดคล้องกับโครงสร้างความสามารถของนักเรียนมากขึ้น ผู้เชี่ยวชาญให้ข้อเสนอแนะและตัดสินใจได้คงเส้นคงวามากขึ้น ช่องว่างระหว่างคะแนนแคบลงและสามารถบรรยายความสามารถนักเรียนในแต่ละระดับได้แม่นยำมากขึ้น

ตัวอย่างงานวิจัยข้างต้นแสดงให้เห็นว่าผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจได้รับประโยชน์จากวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์ เข้าใจและมองเห็นอิทธิพลของปัจจัยต่างๆ ที่ส่งผลต่อคะแนนสอบ สามารถตีความคะแนนสอบและคะแนนจุดตัดได้ เข้าใจความคิดของผู้เชี่ยวชาญคนอื่น มีข้อมูลและหลักฐานความตรงมากพอที่จะอธิบายแนวคิดและข้อเสนอแนะของตนเอง ทำให้สามารถกำหนดคะแนนจุดตัดได้อย่างเหมาะสม ดังนั้นวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์จึงอำนวยความสะดวกให้กับผู้เชี่ยวชาญและผู้มีอำนาจตัดสินใจด้วยสารสนเทศที่เกี่ยวข้องกับข้อสอบและผู้สอบที่เชื่อมโยงกันอย่างเป็นระบบ เป็นวิธีที่ช่วยให้ผลการตัดสินคะแนนจุดตัดคงเส้นคงวา นั้นหมายความว่า เป็นการควบคุมความแปรปรวนของความเห็นจากผู้มีอำนาจตัดสินใจ ทำให้คะแนนจุดตัดที่เกิดขึ้นมีคุณสมบัติความตรงตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา

บทที่ 3

วิธีดำเนินการวิจัย

เนื้อหาในบทนี้เกี่ยวข้องกับแนวคิดและกระบวนการตรวจสอบความสอดคล้องกลมกลืนของข้อมูลคำตอบข้อสอบระดับชาติขั้นพื้นฐาน (Ordinary National Education Test) หรือข้อสอบ ONET กับโครงสร้างทางทฤษฎีหรือแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์และโมเดลการวัดความสามารถทางวิทยาศาสตร์ ดังแสดงในภาพที่ 3.1



ภาพที่ 3.1 กรอบการดำเนินการวิจัย

ขั้นตอนวิจัยในภาพที่ 3.1 เป็นแนวคิดที่พัฒนาจากกระบวนการเปรียบเทียบความสอดคล้องกลมกลืนระหว่างโมเดลการวัดทางทฤษฎีกับโมเดลเกณฑ์มาตรฐานที่เสนอโดย Mellenbergh (2011) โดยเริ่มต้นจากการสร้างแผนที่ภาวะสันนิษฐาน (construct map) ความสามารถทางวิทยาศาสตร์ ด้วยวิธีเชิงคุณภาพ จากนั้นตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานดังกล่าวด้วยวิธีเชิงปริมาณ เมื่อได้แผนที่ภาวะสันนิษฐานที่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์แล้ว จึงให้คะแนนคำตอบข้อสอบตามแผนที่ภาวะสันนิษฐานข้างต้น แล้วนำไปตรวจสอบความสอดคล้องกลมกลืนกับโมเดลการวัดความสามารถทางวิทยาศาสตร์ โมเดลที่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์และสัมพันธ์กับตรรกะการพัฒนาแบบสอบ ONET มากที่สุดจะถูกเลือกเพื่อตรวจสอบความไม่แปรเปลี่ยนของโมเดลด้วยข้อมูลคำตอบข้อสอบ ONET เชิงประจักษ์อีกชุด จากนั้นจะกำหนดคะแนนจุดตัดขั้นต้นตามโมเดลที่ไม่แปรเปลี่ยนดังกล่าว พร้อมกับสร้างสเกลคะแนนความสามารถทางวิทยาศาสตร์และคำอธิบายระดับความสามารถของนักเรียน ดังนั้นเนื้อหาในบทนี้จึงแบ่งออกเป็นสองตอน ตอนแรกคือแนวคิดของกระบวนการวิจัยประกอบด้วยเนื้อหาย่อยสองส่วนคือ เครื่องมือวิจัยและกลุ่มตัวอย่าง และการออกแบบการวิเคราะห์ข้อมูล และตอนที่สองคือผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ รายละเอียดของเนื้อหาแต่ละตอนมีดังต่อไปนี้

ตอนที่ 1 แนวคิดของกระบวนการวิจัย

เครื่องมือวิจัยและกลุ่มตัวอย่าง

เครื่องมือวิจัย

เครื่องมือวิจัยของโครงการนี้คือกรอบโมเดลการวัดที่ใช้วิเคราะห์ข้อมูลจากฐานข้อมูลคำตอบข้อสอบ ONET เพื่อให้ได้จำนวนตัวอย่างที่เพียงพอสำหรับประมาณค่าพารามิเตอร์ตามทฤษฎีการวัดของราล์ซ กรอบโมเดลการวัดที่ใช้คือโมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์แบบสุ่มจากการวัดแบบหลายมิติ (Multidimensional Random Coefficients Multinomial Logit model: Adams, Wilson, & Wang, 1997) หรือโมเดล MRCML ที่ประกอบด้วยพารามิเตอร์สามกลุ่มคือ ความยากข้อสอบ (และค่าความยากระดับชั้น) ค่าเฉลี่ยความสามารถประชากรและค่าความแปรปรวนความสามารถประชากร ผู้วิจัยใช้โมเดล MRCML พัฒนาโมเดลการวัดตามข้อกำหนดลักษณะแบบสอบ (test specification) และคุณสมบัติข้อสอบ (item properties) โมเดลที่พัฒนาขึ้นแบ่งเป็นสองกลุ่ม โดยแต่ละกลุ่มให้คะแนนคำตอบแบบสองค่า (dichotomous) และแบบหลายค่า (polytomous) ประกอบด้วยโมเดลเอกมิติ (unidimensional model) โมเดลพหุมิติระหว่างข้อสอบ (between-item multidimensional model) และโมเดลพหุมิติภายในข้อสอบ (within-item multidimensional model) โดยโมเดลแต่ละกลุ่มมีจำนวนพารามิเตอร์ที่ต้องประมาณไม่เท่ากัน ทำให้ขนาดกลุ่มตัวอย่างขั้นต่ำของแต่ละโมเดลไม่เท่ากันด้วย

หลักการสำหรับกำหนดขนาดกลุ่มตัวอย่างเพื่อตรวจสอบความสอดคล้องของข้อมูลคำตอบข้อสอบ ONET กับโมเดล MRCML ไม่แตกต่างจากหลักการประมาณค่าพารามิเตอร์ของโมเดลสถิติทั่วไปที่กลุ่มตัวอย่างขนาดใหญ่จะให้ค่าประมาณพารามิเตอร์แม่นยำสูงและมีความคลาดเคลื่อนมาตรฐานต่ำ (Bond & Fox, 2007; Linacre, 1994; Stone & Yumoto, 2004) เมื่อพิจารณาโมเดลการวัดทั้งหมดในโครงการวิจัยนี้พบว่า โมเดลพหุมิติภายในข้อสอบแบบให้คะแนนหลายค่ามีจำนวนพารามิเตอร์มากที่สุดเท่ากับ 78 พารามิเตอร์ ดังนั้นเมื่อกำหนดขนาดกลุ่มตัวอย่างให้กับโมเดลนี้แล้วขนาดกลุ่มตัวอย่างที่ได้ก็จะเพียงพอต่อการประมาณค่าพารามิเตอร์ของโมเดลการวัดอื่นที่มีจำนวนพารามิเตอร์น้อยกว่าด้วย

Linacre (1994) และ Wright และ Tennant (1996) เสนอว่าขนาดกลุ่มตัวอย่าง 100 คนต่อหนึ่งพารามิเตอร์ในโมเดลการวัดของ Rasch จะให้ค่าประมาณที่เสถียรอยู่ในช่วงระหว่าง $\pm .5$ ลอจิต (logit) ที่ความเชื่อมั่น 95% และถ้าขนาดกลุ่มตัวอย่างเพิ่มขึ้นเป็น 200 คนต่อหนึ่งพารามิเตอร์ค่าประมาณที่ได้จะเสถียรภายในช่วง $\pm .5$ ลอจิตที่ระดับความเชื่อมั่น 99% นอกจากนี้โมเดลการวัดที่ให้คะแนนคำตอบหลายค่าจะประกอบด้วยพารามิเตอร์ความยากระดับขั้น (item threshold หรือ item step) แทนที่จะเป็นพารามิเตอร์ความยากของข้อสอบทั้งข้อตามแบบข้อสอบที่ให้คะแนนคำตอบแบบสองค่า จึงทำให้ขนาดกลุ่มตัวอย่างเพิ่มขึ้นด้วย ดังนั้นเพื่อให้ค่าประมาณของพารามิเตอร์ต่างๆ ที่ได้จากโมเดล MRCML เสถียรมากที่สุด ผู้วิจัยจึงกำหนดตัวอย่าง 200 คนต่อหนึ่งพารามิเตอร์นั้นหมายความว่าขนาดกลุ่มตัวอย่างขั้นต่ำของโมเดล MRCML ที่มีจำนวนพารามิเตอร์มากที่สุดเท่ากับ 15,600 คน (78 พารามิเตอร์ $\times$ 200 คน)

แต่อย่างไรก็ตามขนาดกลุ่มตัวอย่างที่ผู้วิจัยใช้จริงมีมากกว่าจำนวนข้างต้น โดยผู้วิจัยกำหนดให้ขนาดกลุ่มตัวอย่างสำหรับตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานมีประมาณ 5% สำหรับตรวจสอบความสอดคล้องกลมกลืนของข้อมูลกับโมเดลมีประมาณ 10% และสำหรับตรวจสอบความตรงข้ามกลุ่ม (cross validation) อีกประมาณ 10% ของจำนวนประชากรนักเรียนที่ทำข้อสอบ ONET วิชาวิทยาศาสตร์ หลักการเพิ่มเติมที่ผู้วิจัยใช้กำหนดขนาดกลุ่มตัวอย่างข้างต้นคือ ประการแรกผู้วิจัยสามารถเข้าถึงฐานข้อมูลประชากรจึงสนใจที่จะใช้ข้อมูลทั้งหมดเกิดประโยชน์อย่างเหมาะสม โดยเฉพาะอย่างยิ่งเมื่อการทดสอบ ONET เป็นการทดสอบที่มีส่วนได้เสียสูง การนำคะแนนสอบ ONET มาใช้กำหนดคะแนนจุดตัดจะทำให้ผลการวิจัยมีผลกระทบต่อผู้ที่เกี่ยวข้องอย่างกว้างขวางเช่นเดียวกัน ดังนั้นผลวิจัยที่ได้จากกลุ่มตัวอย่างขนาดใหญ่จึงสร้างความเชื่อมั่นและการยอมรับจากบุคคลทั่วไป

ประการที่สอง เนื่องจากค่าเฉลี่ยคะแนนสอบ ONET วิชาวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 ในอดีตที่ผ่านมามีค่าต่ำกว่าร้อยละ 50 แสดงว่าข้อสอบ ONET บางข้อยากมากทำให้จำนวนนักเรียนที่ตอบถูกข้อสอบเหล่านี้น้อยเมื่อเปรียบเทียบกับจำนวนนักเรียนที่ตอบผิด การใช้กลุ่มตัวอย่างขนาดใหญ่จะทำให้สัดส่วนนักเรียนที่ตอบถูกจากกลุ่มตัวอย่างใกล้เคียงกับข้อมูลประชากร ประการสุดท้าย เมื่อพิจารณาข้อกำหนดลักษณะแบบสอบ ONET ระหว่างปีการศึกษา

2550 ถึง 2552 พบว่าสัดส่วนข้อสอบในแต่ละมาตรฐานการเรียนรู้และสาระการเรียนรู้แตกต่างกันทั้งภายในปีเดียวกันและข้ามปีการศึกษา ผู้วิจัยจึงเลือกใช้กลุ่มตัวอย่างขนาดใหญ่เพื่อช่วยให้ค่าพารามิเตอร์สะท้อนโครงสร้างความสามารถนักเรียนที่แม่นยำมากขึ้น

ประชากรและกลุ่มตัวอย่าง

ประชากรผู้ให้ข้อมูลของโครงการวิจัยนี้คือ นักเรียนชั้นประถมศึกษาปีที่ 6 ที่เข้าสอบ ONET วิชาวิทยาศาสตร์ ปีการศึกษา 2550 ถึง 2552 ซึ่งรวบรวมเป็นฐานข้อมูลของสถาบันทดสอบทางการศึกษาแห่งชาติ หรือ สทศ. ดังตารางที่ 3.1 กลุ่มตัวอย่างประกอบด้วยสามกลุ่มโดยแต่ละกลุ่มถูกสุ่มอย่างง่ายจากฐานข้อมูลประชากรที่กำจนักเรียนที่มีข้อมูลคลาดเคลื่อนและข้อมูลขาดหายอย่างน้อยหนึ่งข้อออกทั้งหมดและเรียกฐานข้อมูลใหม่นี้ว่า ฐานข้อมูลวิเคราะห ข้อมูลคลาดเคลื่อนคือนักเรียนเลือกตัวเลือกมากกว่าหนึ่งตัวเลือกเมื่อข้อสอบข้อนั้นมีคำตอบที่ถูกต้องเพียงตัวเลือกเดียวและเลือกมากกว่าสองตัวเลือกหรือหนึ่งตัวเลือกเมื่อข้อสอบข้อนั้นมีคำตอบที่ถูกต้องสองตัวเลือก ส่วนข้อมูลขาดหายคือ นักเรียนไม่เลือกตัวเลือกใดเลยในข้อสอบแต่ละข้อ

ตารางที่ 3.1 ประชากรและกลุ่มตัวอย่างนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เข้าสอบ ONET วิชาวิทยาศาสตร์ ปีการศึกษา 2550 ถึง 2552

ประชากรและกลุ่มตัวอย่าง	ปีการศึกษา		
	2550	2551	2552
จำนวนนักเรียนที่เข้าสอบทั้งหมด	961,613	953,115	898,372
จำนวนนักเรียนที่ตอบข้อสอบครบถ้วน	894,274	777,900	877,968
กลุ่มตัวอย่างสำหรับตรวจสอบคุณภาพ แผนที่ภาวะสันนิษฐาน	45,000	39,000	44,000
กลุ่มตัวอย่างสำหรับตรวจสอบความสอดคล้องกลมกลืน ระหว่างข้อมูลคำตอบข้อสอบกับโมเดลการวัด	90,000	78,000	88,000
กลุ่มตัวอย่างสำหรับตรวจสอบความตรงข้ามกลุ่ม	90,000	78,000	88,000

หมายเหตุ: ข้อมูลจาก <http://www.onetresult.niets.or.th/AnnouncementWeb/Notice/FrBasicStat.aspx>
เข้าถึงเมื่อวันที่ 26 ธันวาคม 2557

กลุ่มตัวอย่างแรกมีขนาดประมาณ 5% ของฐานข้อมูลวิเคราะห์ กลุ่มตัวอย่างนี้ใช้ตรวจสอบความสอดคล้องกลมกลืนของข้อมูลคำตอบข้อสอบ ONET กับแผนที่ภาวะสันนิษฐานทางทฤษฎีความสามารถทางวิทยาศาสตร์ที่นักเรียนใช้ทำข้อสอบ ONET นั้นหมายความว่า เป็นการหาความสัมพันธ์ระหว่างโครงสร้างสองส่วนคือโครงสร้างข้อมูลคำตอบข้อสอบกับโครงสร้างความสามารถ

ที่ใช้ทำข้อสอบ และเรียกการตรวจสอบแบบนี้ว่าการตรวจสอบความตรงเชิงโครงสร้าง (Embretson, 2007) และนอกจากนี้คำตอบข้อสอบที่สัมพันธ์กับระดับความสามารถของนักเรียนจะใช้เป็นหลักฐานความตรง (validity evidence) ตามนิยามของมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา (AERA, APA, & NCME, 2014) อีกด้วย

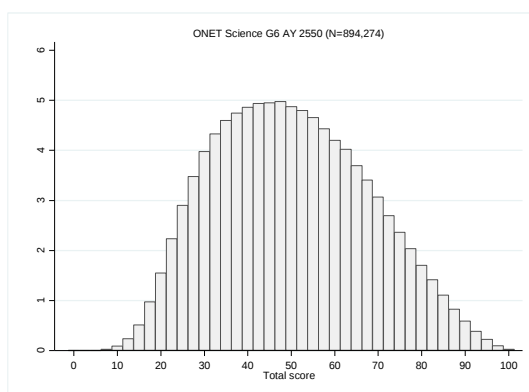
กลุ่มตัวอย่างที่สองมีขนาดประมาณ 10% ของฐานข้อมูลวิเคราะห์ กลุ่มตัวอย่างนี้ใช้ตรวจสอบความสอดคล้องกลมกลืนของข้อคำตอบข้อสอบ ONET กับโมเดลเอมิตี โมเดลพหุมิติ ระหว่างข้อสอบ และโมเดลพหุมิติภายในข้อสอบ ข้อสังเกตเกี่ยวกับความแตกต่างของข้อมูลกลุ่มตัวอย่างแรกกับกลุ่มตัวอย่างที่สองคือ เมื่อได้แผนที่ภาวะสันนิษฐานที่สัมพันธ์กับโครงสร้างข้อมูลคำตอบข้อสอบแล้ว จะใช้แผนที่ภาวะสันนิษฐานดังกล่าวเป็นแนวทางให้คะแนนคำตอบข้อสอบ ONET ใหม่ซึ่งแนวทางดังกล่าวจะแตกต่างจากวิธีให้คะแนนดั้งเดิมของ สทศ. คือแทนที่จะแบ่งความสามารถของนักเรียนออกเป็นกลุ่มนักเรียนที่ตอบข้อสอบนั้นๆ ถูกและกลุ่มตอบผิดแล้วให้คะแนนเพียงสองค่าคือได้คะแนนเต็มของข้อนั้นกับไม่ได้คะแนนเลย และอาจจะละเลยความสัมพันธ์ระหว่างความยากข้อสอบกับระดับความสามารถของนักเรียนที่ใช้ทำข้อสอบข้อนั้น เป็นการให้คะแนนที่คำนึงถึงโครงสร้างความสามารถที่นักเรียนใช้ทำข้อสอบและระดับความยากข้อสอบที่แตกต่างกัน คะแนนคำตอบข้อสอบที่ได้จากแนวทางให้คะแนนแบบใหม่นี้จะเป็นมาตรวัดแบบจัดอันดับ (ordinal scale) ที่จัดกลุ่มนักเรียนตามระดับความสามารถที่ใช้แก้โจทย์ปัญหาที่มีระดับความยากแตกต่างกัน

กลุ่มตัวอย่างสุดท้ายมีขนาด 10% ของฐานข้อมูลวิเคราะห์ กลุ่มตัวอย่างนี้ใช้ตรวจสอบว่าผลวิเคราะห์และค่าพารามิเตอร์ที่ประมาณได้จากโมเดลการวัดกับกลุ่มตัวอย่างที่สองจะมีสมบัติแตกต่างจากเดิมหรือไม่ เมื่อใช้ข้อมูลจากกลุ่มตัวอย่างอีกกลุ่มหนึ่งที่มาจากประชากรเดียวกัน ดังนั้นจึงต้องตรวจสอบโมเดลที่สอดคล้องกลมกลืนที่สุดก่อน แล้วจึงตรวจสอบว่าโมเดลที่สอดคล้องกลมกลืนที่สุดนั้นมีสมบัติความตรงข้ามกลุ่มหรือไม่

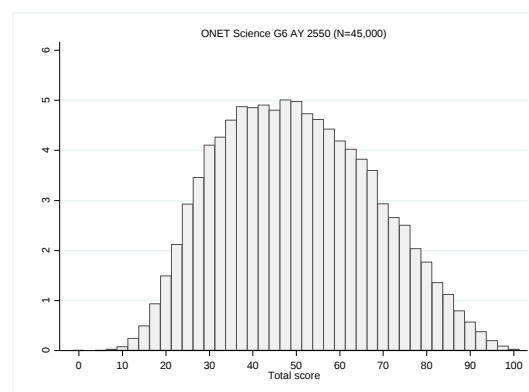
ตารางที่ 3.2 ถึง 3.4 แสดงให้เห็นว่าค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET ของฐานข้อมูลวิเคราะห์และกลุ่มตัวอย่างชุดต่างๆ มีลักษณะคล้ายคลึงกับฐานข้อมูลประชากรนักเรียนที่เข้าสอบทั้งหมด ดังนั้นจึงสามารถสรุปได้ว่าข้อมูลที่ได้จากตัวอย่างทั้งสามกลุ่มเป็นตัวแทนข้อมูลจากประชากรนักเรียน และเป็นหลักฐานยืนยันว่าเมื่อไม่นำข้อมูลของนักเรียนที่ตอบไม่สมบูรณ์และคลาดเคลื่อนมาใช้วิเคราะห์จะไม่ส่งผลให้ผลวิจัยคลาดเคลื่อนและลำเอียง

ตารางที่ 3.2 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์
ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2550

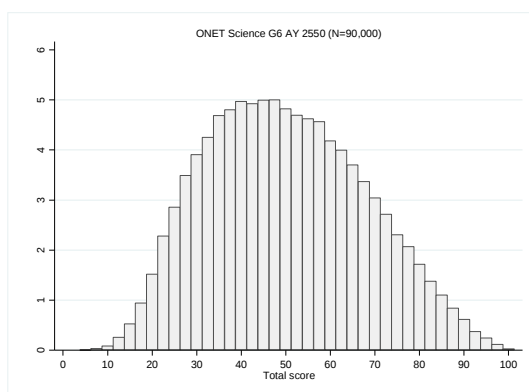
ประชากรและกลุ่มตัวอย่าง	ค่าสถิติพื้นฐาน ¹						
	N	Mean	SD	Min	Max	Median	Mode
ประชากรนักเรียนที่เข้าสอบทั้งหมด	961,613	49.57	17.70	0	100	47.50	47.50
ประชากรนักเรียนที่ตอบครบถ้วน (a)	894,274	50.07	17.70	0	100	50	47.50
กลุ่มตัวอย่างสำหรับตรวจสอบ แผนที่ภาวะสัณนิษฐาน (b)	45,000	50.13	17.63	0	100	50	47.50
กลุ่มตัวอย่างสำหรับตรวจสอบ ความสอดคล้องกลมกลืน (c)	90,000	50.08	17.70	5	100	50	45, 47.50 ²
กลุ่มตัวอย่างสำหรับตรวจสอบ ความตรงข้ามกลุ่ม (d)	90,000	50.12	17.65	2.50	100	50	47.50



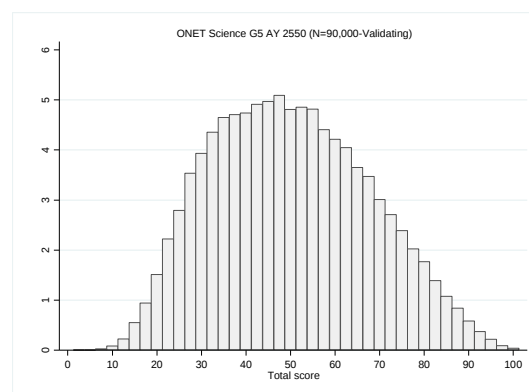
(a)



(b)



(c)



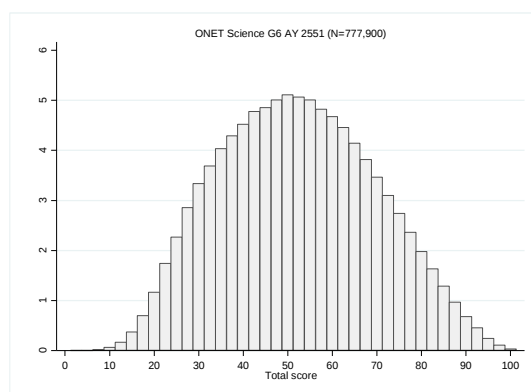
(d)

หมายเหตุ ¹ ค่าสถิติพื้นฐานของคะแนนดิบที่มีคะแนนเต็มเท่ากับ 100 คะแนน จากข้อสอบ 40 ข้อ ข้อละ 2.5 คะแนน

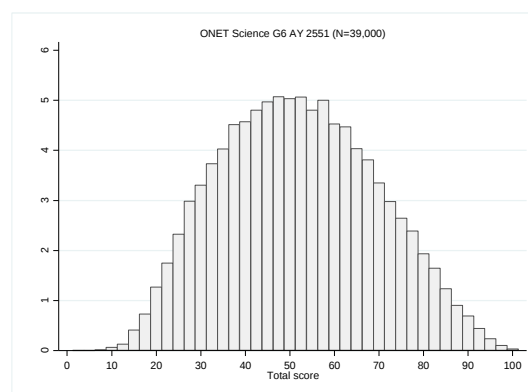
² มีความถี่สูงสุดสองค่า

ตารางที่ 3.3 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์
ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2551

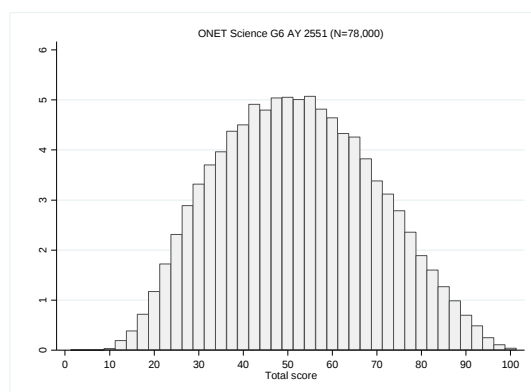
ประชากรและกลุ่มตัวอย่าง	ค่าสถิติพื้นฐาน ¹						
	N	Mean	SD	Min	Max	Median	Mode
ประชากรนักเรียนที่เข้าสอบทั้งหมด	953,155	51.68	17.28	0	100	50	50
ประชากรนักเรียนที่ตอบครบถ้วน (a)	777,900	52.23	17.47	2.50	100	52.50	50
กลุ่มตัวอย่างสำหรับตรวจสอบ แผนที่ภาวะสันนิษฐาน (b)	39,000	51.95	17.47	2.50	100	52.50	47.50, 52.50 ²
กลุ่มตัวอย่างสำหรับตรวจสอบ ความสอดคล้องกลมกลืน (c)	78,000	52.20	17.49	2.50	100	52.50	55
กลุ่มตัวอย่างสำหรับตรวจสอบ ความตรงข้ามกลุ่ม (d)	78,000	52.27	17.53	2.50	100	52.50	52.50



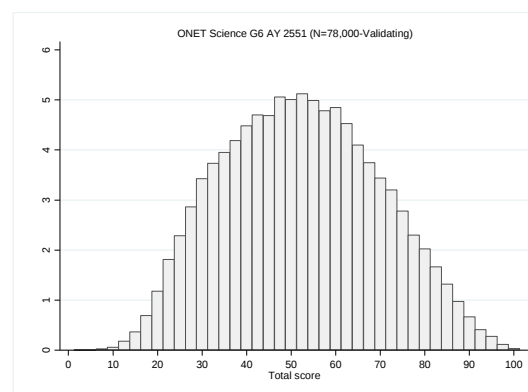
(a)



(b)



(c)



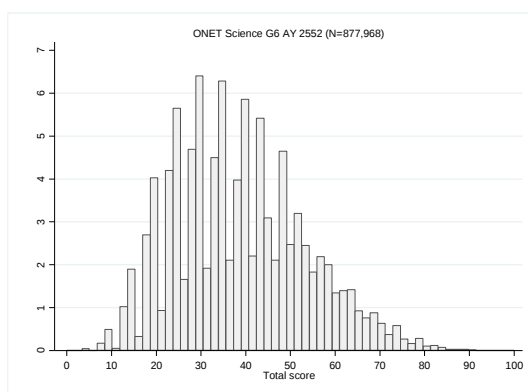
(d)

หมายเหตุ ¹ ค่าสถิติพื้นฐานของคะแนนดิบที่มีคะแนนเต็มเท่ากับ 100 คะแนน จากข้อสอบ 40 ข้อ ข้อละ 2.5 คะแนน

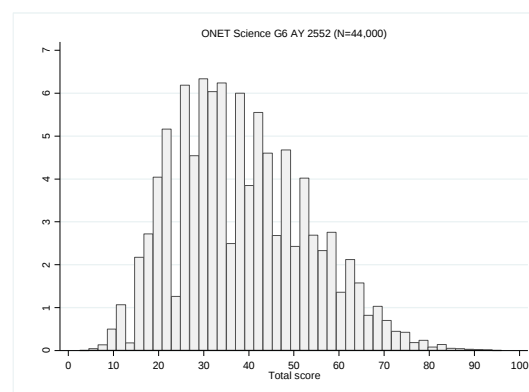
² มีความถี่สูงสุดสองค่า

ตารางที่ 3.4 ค่าสถิติพื้นฐานและการแจกแจงคะแนนสอบ ONET วิชาวิทยาศาสตร์
ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2552

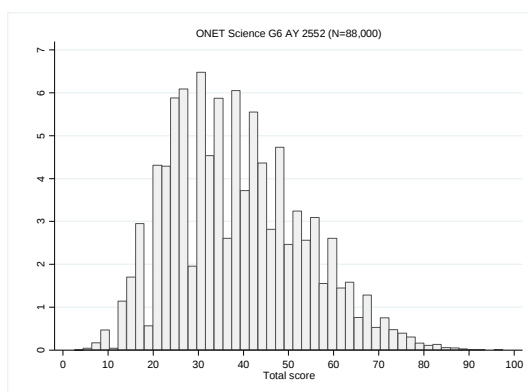
ประชากรและกลุ่มตัวอย่าง	ค่าสถิติพื้นฐาน ¹						
	N	Mean	SD	Min	Max	Median	Mode
ประชากรนักเรียนที่เข้าสอบทั้งหมด	898,372	38.67	14.72	0	100	37.50	27.50
ประชากรนักเรียนที่ตอบครบถ้วน (a)	877,968	38.68	14.73	0	100	37.50	27.50
กลุ่มตัวอย่างสำหรับตรวจสอบ แผนที่ภาวะสัณนิฐาน (b)	44,000	38.61	14.67	0	96	37.50	25
กลุ่มตัวอย่างสำหรับตรวจสอบ ความสอดคล้องกลมกลืน (c)	88,000	38.72	14.71	2.50	97.50	37.50	27.50
กลุ่มตัวอย่างสำหรับตรวจสอบ ความตรงข้ามกลุ่ม (d)	88,000	38.68	14.76	2.50	100	37.50	27.50



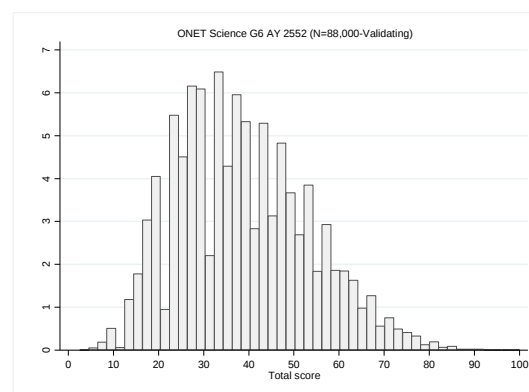
(a)



(b)



(c)



(d)

หมายเหตุ ¹ ค่าสถิติพื้นฐานของคะแนนดิบที่มีคะแนนเต็มเท่ากับ 100 คะแนน จากข้อสอบ 37 ข้อ ข้อ 1-32 ข้อละ

2.5 คะแนน ข้อ 33-37 ข้อละ 4 คะแนน

การออกแบบการวิเคราะห์ข้อมูล

ข้อกำหนดคุณลักษณะแบบสอบ คุณสมบัติข้อสอบและระดับความยากข้อสอบ ONET

การออกแบบการวิเคราะห์ข้อมูลของโครงการวิจัยนี้เป็นการแปลงกรอบแนวคิดของการวิเคราะห์คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ (ดูภาพที่ 1.1 หน้า 13) ให้อยู่ในรูปของขั้นตอนวิธีในภาพที่ 3.1 กระบวนการวิจัยเริ่มต้นด้วยการสร้างแผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากผลวิเคราะห์ข้อมูลโดยผู้เชี่ยวชาญเกี่ยวกับโครงสร้างความสามารถทางวิทยาศาสตร์ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ได้แก่ สารการเรียนรู้ มาตรฐานการเรียนรู้ ตัวชี้วัดและคำบรรยายที่เกี่ยวข้อง และผลวิเคราะห์เกี่ยวกับคุณสมบัติข้อสอบ ONET ได้แก่ ลักษณะโจทย์คำถามและตัวเลือก ระดับความยากหรือความซับซ้อนของข้อสอบและเป้าหมายผลการเรียนรู้ที่ข้อสอบมุ่งวัด ซึ่งโดยปกติแล้วแผนที่ภาวะสันนิษฐานหรือโมเดลทักษะการคิด (cognition model) ควรสร้างขึ้นก่อนการสร้างข้อสอบ (Brown & Wilson, 2011; Mislevy, 1996; National Research Council, 2001; Wilson, 2005) เช่นเดียวกับการสร้างคำบรรยายระดับความสามารถ (performance level descriptor) ในกระบวนการกำหนดคะแนนจุดตัดที่ควรสร้างก่อนกำหนดคะแนนจุดตัด และควรจะใช้คำบรรยายระดับความสามารถดังกล่าวชี้แนะแนวทางสร้างข้อสอบและใช้พัฒนาข้อกำหนดลักษณะแบบสอบด้วย (Bejar, Braun, & Tannenbaum, 2007; Perie, 2008) สำหรับโครงการวิจัยนี้ เมื่อพิจารณากระบวนการสร้างแบบสอบ ONET พบว่าไม่มีโมเดลความสามารถทางวิทยาศาสตร์ปรากฏอย่างชัดเจน จึงจำเป็นต้องนำผลวิเคราะห์ดังกล่าวข้างต้นมาใช้สร้างโมเดลความสามารถทางวิทยาศาสตร์ร่วมกับข้อกำหนดลักษณะแบบสอบ ONET ที่ได้รับจาก สทศ.

ผู้เชี่ยวชาญของโครงการวิจัยนี้คือครูจำนวน 28 คน โดยผู้เชี่ยวชาญแต่ละคนมีประสบการณ์การสอนวิชาวิทยาศาสตร์อย่างน้อย 5 ปี แบ่งออกเป็นระดับประถมศึกษาตอนปลายจำนวน 10 คน และระดับมัธยมศึกษาตอนต้นจำนวน 18 คน และเคยเข้าร่วมอบรมด้านการวัดและการประเมินการเรียนรู้อย่างน้อย 2 ครั้งในช่วงเวลา 3 ปีที่ผ่านมา นอกจากนี้ผู้เชี่ยวชาญทุกคนยังได้รับฟังคำชี้แจงเกี่ยวกับวิธีวิเคราะห์ข้อสอบ ONET ตามแนวทางของโครงการวิจัยนี้ก่อนเริ่มวิเคราะห์จริงด้วย

ตารางที่ 3.5 แสดงข้อกำหนดลักษณะแบบสอบ ONET ปีการศึกษา 2550 ถึง 2552 โดยผู้เชี่ยวชาญได้วิเคราะห์ข้อมูลเพิ่มเติมจากข้อกำหนดลักษณะแบบสอบที่ได้รับจาก สทศ. ดังนี้ ประการแรก เนื่องจากข้อสอบ ONET ปีการศึกษา 2550 สร้างตามหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544 (กรมวิชาการ, 2545) แต่ข้อสอบ ONET ปีการศึกษา 2551-2552 สร้างตามหลักสูตรแกนกลางฯ พ.ศ. 2551 ดังนั้นเพื่อให้ข้อสอบทั้งสามปีสะท้อนโครงสร้างความสามารถทางวิทยาศาสตร์เดียวกัน ผู้เชี่ยวชาญจึงวิเคราะห์ข้อสอบ ONET และระบุข้อกำหนดลักษณะแบบสอบ ONET ปีการศึกษา 2550 ใหม่ให้สอดคล้องกับหลักสูตรแกนกลางฯ พ.ศ. 2551 ดังแสดงในตารางที่ 3.5

ตารางที่ 3.5 ข้อกำหนดลักษณะแบบสอบ ONET วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6

สาระการเรียนรู้ ตามหลักสูตรแกนกลางฯ	มาตรฐาน การเรียนรู้	ปีการศึกษา 2550 ¹		ปีการศึกษา 2551		ปีการศึกษา 2552	
		จำนวน	ข้อที่	จำนวน	ข้อที่	จำนวน	ข้อที่
สาระที่ 1: สิ่งมีชีวิตกับกระบวนการ ดำรงชีวิต	ว 1.1	6	1, 2, 3, 4, 5, 6	3	1, 3, 4	7	2, 3, 4, 25, 26, 27, 37
	ว 1.2	1	7	4	2, 5, 6, 7	2	5, 6
สาระที่ 2: ชีวิตกับสิ่งแวดล้อม	ว 2.1	2	8, 9	3	8, 9, 10	1	33
	ว 2.2	-	-	-	-	1	1
สาระที่ 3: สารและสมบัติ ของสาร	ว 3.1	8	10, 11, 12, 13, 14, 15, 16, 17	7	11, 12, 13, 14, 15, 16, 17	8	8, 9, 10, 11, 12, 13, 15, 16
	ว 3.2	2	18, 19	1	18	-	-
สาระที่ 4: แรงและการเคลื่อนที่	ว 4.1	2	20, 21	2	19, 20	4	7, 14, 17, 34
	ว 4.2	1	22	1	21	1	18
สาระที่ 5: พลังงาน	ว 5.1	7	23, 24, 25, 26, 27, 28, 29	6	22, 23, 24, 25, 26, 27	5	19, 20, 21, 22, 35
สาระที่ 6: กระบวนการ เปลี่ยนแปลงของโลก	ว 6.1	6	30, 31, 32, 33, 34, 35	6	28, 29, 30, 31, 32, 33	3	23, 24, 30
สาระที่ 7: ดาราศาสตร์และอวกาศ	ว 7.1	4	37, 38, 39 40	4	34, 35, 36, 37	5	28, 29, 31, 32, 36
	ว 7.2	1	36	3	38, 39, 40	-	-
สาระที่ 8: ธรรมชาติของวิทยาศาสตร์ และเทคโนโลยี	ว 8.1	(5) ²	8, 10, 26, 31, 33	(9) ²	3, 14, 18, 21, 24, 27, 28, 30, 31	(8) ²	3, 4, 7, 10, 11, 13, 15, 27
รวม		40		40	-	37	-

หมายเหตุ ¹ แก่ไขมาตรฐานการเรียนรู้ที่วัดโดยข้อสอบ ONET ปีการศึกษา 2550 จากหลักสูตรการศึกษาขั้นพื้นฐาน

พ.ศ. 2548 เป็นหลักสูตรแกนกลางฯ พ.ศ. 2551

² ผลประเมินโดยผู้เชี่ยวชาญที่มีความเห็นว่าข้อสอบเหล่านั้นนอกจากจะวัดมาตรฐานอื่นตามที่ระบุไว้ข้างต้นแล้วยังวัดมาตรฐาน ว 8.1 ด้วย ดังนั้นผลรวมจำนวนข้อสอบทั้งฉบับจึงไม่รวมข้อสอบที่วัดมาตรฐาน ว 8.1

ประการที่สอง เมื่อพิจารณาข้อกำหนดลักษณะแบบสอบที่ได้รับจาก สทศ. พบว่าไม่มีข้อสอบใดเลยที่ระบุโดยตรงว่าวัดมาตรฐาน ว 8.1 (เกี่ยวกับทักษะกระบวนการทางวิทยาศาสตร์) แต่ถือว่าสาระของมาตรฐาน ว 8.1 แทรกอยู่ในมาตรฐานอื่นๆ โดยแต่ละมาตรฐานประกอบด้วยมาตรฐานเนื้อหาและมาตรฐานการปฏิบัติ (ดูตารางที่ A1 ในภาคผนวก) จึงทำให้เข้าใจได้ว่าข้อสอบ ONET ทุกข้อวัดมาตรฐาน ว 8.1 ด้วย แต่อย่างไรก็ตามข้ออ้างดังกล่าวอาจนำไปสู่ปัญหาการระบุ

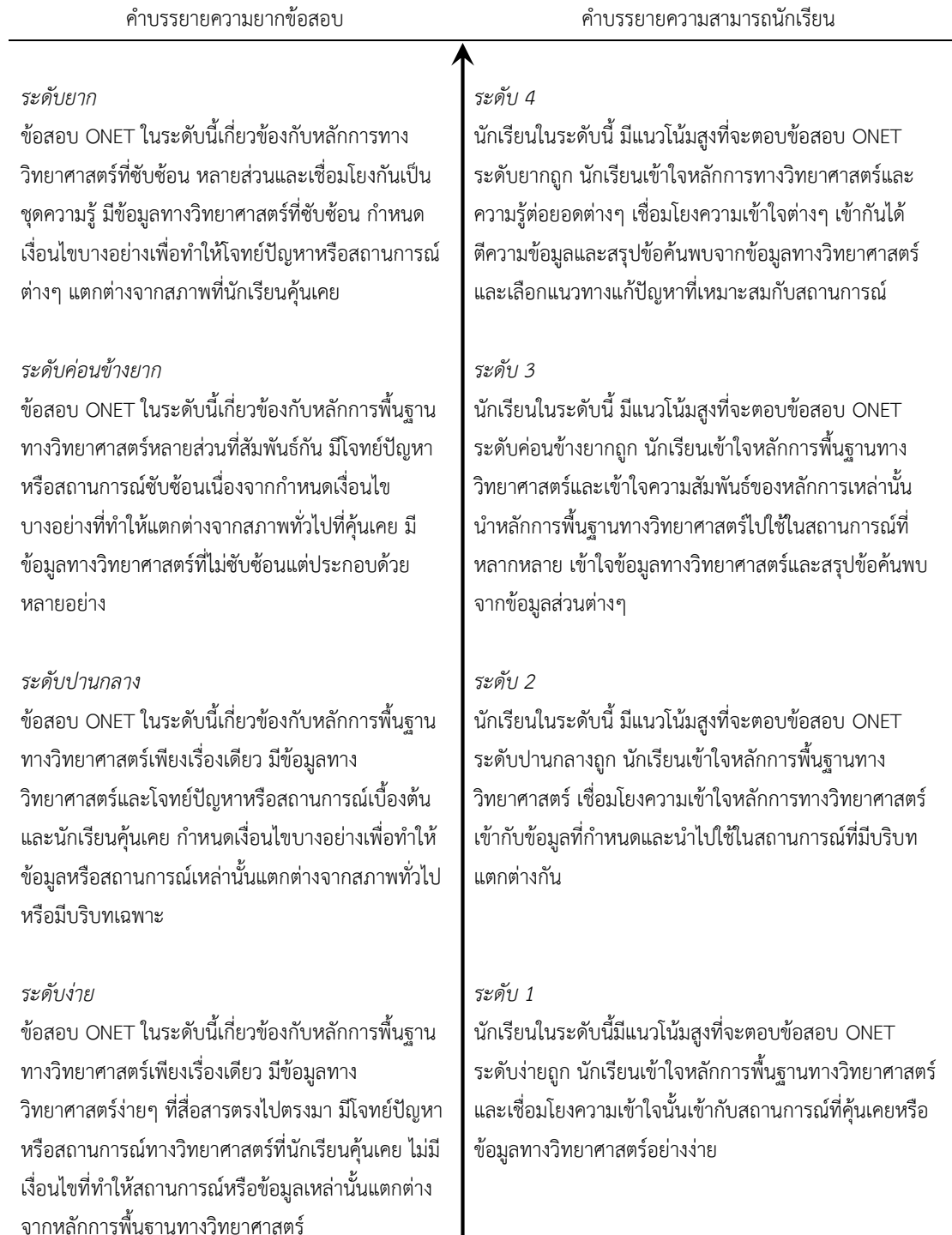
ข้อกำหนดลักษณะโมเดลคลาดเคลื่อน (model misspecification) ดังนั้นผู้เชี่ยวชาญจึงวิเคราะห์คุณสมบัติข้อสอบ ONET ของแต่ละปีการศึกษาอีกครั้งแล้วประเมินว่าข้อสอบข้อใดบ้างที่วัดมาตรฐาน 8.1 โดยข้อสอบแต่ละข้อจะถูกประเมินโดยผู้เชี่ยวชาญระหว่าง 3-7 คน เมื่อพิจารณาดัชนีความคล้ายคลึงของผลประเมิน (similarity index) ซึ่งคำนวณจากจำนวนผู้เชี่ยวชาญที่มีความเห็นตรงกันต่อจำนวนผู้เชี่ยวชาญทั้งหมด (ดูตารางที่ A2 ในภาคผนวก) และกำหนดเกณฑ์ว่าข้อสอบที่มีค่าดัชนีดังกล่าวมากกว่า .5 เป็นข้อสอบที่วัดมาตรฐาน 8.1 ด้วย จากตารางที่ 3.5 แสดงว่าจำนวนข้อสอบที่วัดมาตรฐาน 8.1 ของแต่ละปีการศึกษาไม่เท่ากัน และกระจายอยู่ในมาตรฐานต่างๆ ข้อค้นพบนี้ช่วยยืนยันเบื้องต้นว่าข้อสอบ ONET ทุกข้ออาจจะไม่ได้วัดมาตรฐาน 8.1

ประการที่สาม เมื่อพิจารณาตัวชี้วัดของแต่ละมาตรฐานพบว่า ข้อสอบ ONET สามารถระบุให้ชัดเจนมากขึ้นว่ามีเนื้อหาสัมพันธ์กับตัวชี้วัดใดตามหลักสูตรแกนกลางฯ พ.ศ. 2551 (สำนักวิชาการและมาตรฐานการศึกษา, 2551) โดยแต่ละมาตรฐานจะประกอบด้วยชุดตัวชี้วัดที่จำแนกตามระดับชั้น ตั้งแต่ประถมศึกษาปีที่ 4 ถึง 6 และบางมาตรฐานก็ไม่มีตัวชี้วัดครบทุกระดับชั้น ดังนั้นเพื่อให้การวิเคราะห์ข้อสอบ ONET ได้สารสนเทศมากที่สุดและเป็นประโยชน์ต่อการสร้างแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ ผู้เชี่ยวชาญจึงประเมินว่าข้อสอบ ONET แต่ละข้อนั้นวัดตัวชี้วัดใดด้วย ค่าดัชนีความคล้ายคลึงที่มากกว่า .5 จะบ่งชี้ว่าข้อสอบข้อนั้นสอดคล้องกับตัวชี้วัดที่กำหนด สำหรับข้อสอบบางข้อที่วัดมาตรฐาน 8.1 ร่วมกับมาตรฐานอื่นๆ ผู้เชี่ยวชาญไม่ได้วิเคราะห์ว่าข้อสอบข้อนั้นสัมพันธ์กับตัวชี้วัดใดของมาตรฐาน 8.1 เนื่องจากต้องการสะท้อนภาพรวมของข้อสอบที่สัมพันธ์กับมาตรฐาน 8.1 มากกว่า ผลประเมินความสอดคล้องระหว่างความสามารถนักเรียนใช้ทำข้อสอบกับตัวชี้วัดแสดงให้เห็นว่าข้อสอบแต่ละข้อสัมพันธ์กับตัวชี้วัดอย่างน้อย 1 ตัวและผู้เชี่ยวชาญมีความเห็นในทิศทางเดียวกัน (ดูรายละเอียดในตารางที่ A3 ในภาคผนวก)

ประการสุดท้าย ธรรมชาติของข้อสอบ ONET มีระดับความยากแตกต่างกัน นักเรียนที่ตอบถูกข้อยากได้นั้นต้องใช้ความสามารถมากกว่าตอบถูกข้อง่าย ดังนั้นระดับความยากข้อสอบจึงสัมพันธ์กับระดับความสามารถที่นักเรียนใช้ทำข้อสอบ นักเรียนที่ตอบถูกข้อยากมีแนวโน้มหรือความน่าจะเป็นสูงมากที่จะมีความสามารถสูงกว่านักเรียนที่ตอบถูกเฉพาะข้อง่าย ในทำนองเดียวกันนักเรียนที่ตอบถูกเฉพาะข้อง่ายก็มีแนวโน้มสูงมากที่จะมีความสามารถน้อยกว่านักเรียนที่ตอบถูกข้อยาก แม้ว่าในบางกรณีแนวโน้มดังกล่าวอาจจะแปรปรวน แต่ก็สามารถบรรยายปรากฏการณ์ในภาพรวมได้เป็นอย่างดี ดังนั้นสำหรับโครงการวิจัยนี้ ผู้เชี่ยวชาญจึงวิเคราะห์ระดับความยากข้อสอบโดยพิจารณาจากคุณสมบัติข้อสอบว่ามีความซับซ้อนมากน้อยเพียงใด และความสามารถนักเรียนที่ใช้แก้ปัญหาข้อสอบในแต่ละระดับความซับซ้อนว่ามีลักษณะอย่างไรด้วย

ภาพที่ 3.2 แสดงถึงระดับความยากข้อสอบ ONET ที่ผู้เชี่ยวชาญได้วิเคราะห์และบรรยายเชื่อมโยงกับความสามารถที่นักเรียนใช้ทำข้อสอบแบ่งเป็นสี่ระดับ เริ่มตั้งแต่ระดับง่าย (ด้านล่างสุดของภาพ) สู่ระดับยาก (ด้านบนสุดของภาพ) ลูกศรชี้ขึ้นจากด้านล่างสู่ด้านบนแสดงถึงระดับความยากข้อสอบที่สะสมเพิ่มขึ้นและระดับความสามารถที่ใช้ทำข้อสอบก็จะเพิ่มมากขึ้นเมื่อข้อสอบยากขึ้น เส้น

ลูกศรนี้จึงทำหน้าที่เป็นสเกลที่บอกตำแหน่ง (location) ของนักเรียนในแต่ละกลุ่มที่สามารถตอบถูกข้อสอบที่ยากง่ายแตกต่างกัน โครงสร้างความยากข้อสอบและความสามารถของนักเรียนที่ใช้ทำข้อสอบนี้มีแนวคิดเช่นเดียวกับสเกลเชิงสะสม (cumulative scale) ตามแนวคิดของ Guttman (1944, 1950, 1954) โดยปริมาณคุณลักษณะจะมีลักษณะเดียวกับโครงสร้างคุณลักษณะนั้นที่สะสม



ภาพที่ 3.2 โครงสร้างความยากข้อสอบ ONET และความสามารถทางวิทยาศาสตร์

เพิ่มขึ้นจากระดับหนึ่งสู่อีกระดับหนึ่ง ผลประเณนเกี่ยวกับระดับความยากข้อสอบพบว่า ผู้เชี่ยวชาญส่วนใหญ่ประเมินระดับความยากข้อสอบแต่ละข้อเป็นไปในทิศทางเดียวกัน (ดูรายละเอียดในตารางที่ A4 ในภาคผนวก)

จากภาพที่ 3.2 จะเห็นว่าข้อสอบระดับง่ายคือข้อสอบที่มีโจทย์คำถามและตัวเลือกไม่ซับซ้อน ไม่มีเงื่อนไขที่ทำให้โจทย์ปัญหาหรือสถานการณ์แตกต่างจากหลักการพื้นฐานทางวิทยาศาสตร์ที่นักเรียนคุ้นเคย รวมทั้งอาจมีข้อมูลทางวิทยาศาสตร์อย่างง่ายที่สื่อสารตรงไปตรงมาเกี่ยวข้องด้วย นักเรียนที่ตอบถูกข้อสอบระดับง่ายจึงใช้เพียงความรู้พื้นฐานทางวิทยาศาสตร์และถือว่าเป็นความสามารถระดับแรก นักเรียนไม่จำเป็นต้องวิเคราะห์หรือเชื่อมโยงหลักการพื้นฐานหลายอย่างเพื่อแก้โจทย์ปัญหาหรือสถานการณ์และไม่จำเป็นต้องตีความข้อมูลเพื่อสรุปเป็นข้อค้นพบด้วย

ในทางตรงกันข้าม ข้อสอบระดับยากหมายถึงข้อสอบที่มีโจทย์ปัญหาและตัวเลือกซับซ้อน เนื้อหาข้อสอบเกี่ยวข้องกับหลักการทางวิทยาศาสตร์หลายอย่าง ทั้งหลักการพื้นฐานและหลักการเพิ่มเติม สถานการณ์ที่กำหนดขึ้นซับซ้อนไม่ว่าจะเป็นการกำหนดเงื่อนไขบางอย่างที่ทำให้มีสภาพแตกต่างจากสภาพทั่วไปหรือการสร้างสถานการณ์ใหม่ที่นักเรียนไม่คุ้นเคย ส่วนข้อมูลทางวิทยาศาสตร์ที่กำหนดขึ้นก็ซับซ้อนและต้องตีความข้อมูลเหล่านั้นเพื่อสรุปเป็นข้อค้นพบ นักเรียนที่ตอบถูกข้อสอบระดับยากนี้จึงต้องเข้าใจหลักการพื้นฐานทางวิทยาศาสตร์และต้องนำหลักการเหล่านั้นมาใช้อย่างเหมาะสม นักเรียนจะต้องตีความข้อมูลและสรุปเป็นสาระสำคัญเพื่อจะได้เชื่อมโยงเข้ากับหลักการพื้นฐานทางวิทยาศาสตร์ต่างๆ ได้อย่างเหมาะสม

ดังนั้นเมื่อพิจารณาโครงสร้างความยากข้อสอบ ONET จะพบว่า ระดับความยากข้อสอบที่เพิ่มสูงขึ้นมีความหมายไม่เพียงแค่ว่าปริมาณความรู้หรือหลักการทางวิทยาศาสตร์ที่ต้องใช้แก้ไขโจทย์ปัญหาเท่านั้น แต่รวมถึงความซับซ้อนของหลักการที่ใช้ ความคุ้นเคยของนักเรียนที่มีต่อสถานการณ์ ความซับซ้อนของสถานการณ์ ความสัมพันธ์ระหว่างสถานการณ์หรือโจทย์ปัญหาและหลักการทางวิทยาศาสตร์ และความซับซ้อนของข้อมูลทางวิทยาศาสตร์ที่กำหนดขึ้นด้วย ซึ่งปัจจัยเหล่านี้สะท้อนให้เห็นว่านักเรียนต้องใช้ทักษะการคิดที่หลากหลายและซับซ้อนมากขึ้นเมื่อข้อสอบยากเพิ่มขึ้น

แผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์

เมื่อพิจารณาข้อกำหนดลักษณะแบบสอบ ONET ตั้งแต่ปีการศึกษา 2550 ถึง 2552 พบว่า จำนวนข้อสอบที่วัดแต่ละมาตรฐานการเรียนรู้ไม่เท่ากัน เช่นจำนวนข้อสอบที่วัดมาตรฐาน ว 1.1 ในปีการศึกษา 2551 มีประมาณครึ่งหนึ่งของข้อสอบปีการศึกษา 2550 และ 2552 ในขณะที่มาตรฐาน ว 1.2 กลับมีจำนวนข้อสอบมากกว่าสองเท่าของข้อสอบปีการศึกษา 2550 และ 2552 หรือในกรณีมาตรฐาน ว 2.2 พบว่า ไม่มีข้อสอบใดเลยที่วัดมาตรฐานนี้ยกเว้นข้อสอบปีการศึกษา 2552 ที่มีจำนวนเพียงข้อเดียว ในทางตรงกันข้ามมาตรฐาน 3.2 กลับไม่มีข้อสอบใดเลยวัดมาตรฐานนี้ทั้งๆ ที่เมื่อปีการศึกษา 2550 และ 2551 มีข้อสอบวัดมาตรฐานนี้

ข้อค้นพบเกี่ยวกับความไม่คงที่ของจำนวนข้อสอบที่วัดแต่ละมาตรฐานการเรียนรู้ แสดงให้เห็นว่าการสร้างแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์รายมาตรฐานการเรียนรู้ไม่เหมาะสม กล่าวคือ ประการแรก แผนที่ภาวะสันนิษฐานที่พัฒนาขึ้นจะไม่สามารถบรรยายและสรุปอ้างอิงแนวโน้มความสามารถของนักเรียนที่เปลี่ยนแปลงข้ามปีการศึกษาได้ ประการที่สองคือ แผนที่ภาวะสันนิษฐานที่เกิดขึ้นจากข้อสอบจำนวนน้อยโดยเฉพาะมาตรฐานการเรียนรู้ที่ถูกวัดโดยข้อสอบเพียงข้อเดียวในแต่ละปีการศึกษา จะขาดความตรงเชิงเนื้อหา ไม่มีความตรงเชิงโครงสร้าง ข้อสอบกับแผนที่ภาวะสันนิษฐานไม่สอดคล้องกลมกลืน และค่าประมาณพารามิเตอร์ไม่แม่นยำ แม้ว่าในขั้นตอนสร้างข้อสอบจะสามารถระบุได้อย่างชัดเจนว่าข้อสอบข้อนั้นวัดมาตรฐานการเรียนรู้ใด แต่ข้อสอบก็ขาดสมบัติความเป็นตัวแทนของมวลเนื้อหาความรู้ที่มุ่งวัด ความเป็นตัวแทนของประชากรข้อสอบ และความเป็นตัวแทนของคำตอบถูกและคำตอบผิดที่มีต่อมาตรฐานการเรียนรู้นั้น ในกรณีเช่นนี้ Wright (1992) ถือว่าขาดความตรงเชิงเนื้อหา เพราะเนื้อหาของมาตรฐานการเรียนรู้ที่วัดด้วยข้อสอบเพียงข้อเดียวนั้นไม่ถูกตรวจสอบด้วยข้อสอบข้ออื่นที่วัดมาตรฐานการเรียนรู้เดียวกัน นั่นหมายความว่าแผนที่ภาวะสันนิษฐานที่สร้างจากข้อสอบเพียงไม่กี่ข้อจะขาดความตรงเชิงเนื้อหาด้วย

ข้อสอบข้อเดียวไม่สามารถนิยามโครงสร้างความสามารถนักเรียนของแต่ละมาตรฐานเพราะนักเรียนจะถูกแบ่งออกเป็นเพียงสองกลุ่มคร่าวๆ คือกลุ่มที่ตอบถูกและตอบผิดข้อสอบข้อนั้น และภายในแต่ละกลุ่มก็มีความผันแปรด้วยทั้งนี้ขึ้นอยู่กับความยากของข้อสอบข้อนั้น ในกรณีเช่นนี้ Wright (1992) เรียกว่าขาดความตรงเชิงโครงสร้าง เพราะกลุ่มนักเรียนที่ตอบถูกและตอบผิดข้อสอบเพียงข้อเดียวให้สารสนเทศไม่เพียงพอเพื่อบรรยายโครงสร้างความสามารถนักเรียน นั่นหมายความว่าแผนที่ภาวะสันนิษฐานที่สร้างจากข้อสอบเพียงข้อเดียวจะไม่สะท้อนโครงสร้างความสามารถทางวิทยาศาสตร์ของนักเรียนด้วย

นอกจากนี้ข้อสอบเพียงข้อเดียว (หรือสามข้อก็ตาม) ไม่ให้ข้อมูลที่แม่นยำมากพอเกี่ยวกับความสามารถของนักเรียน (Wright, 1992) นักเรียนที่ตอบถูกข้อสอบเพียงข้อเดียวไม่สามารถสรุปได้ว่านักเรียนคนนั้นมีความรู้ความเข้าใจเป็นอย่างดี เนื่องจากนักเรียนอาจจะเดาถูกหรือตัดตัวเลือกที่น่าจะผิดได้ นั่นหมายความว่าคำตอบถูกและคำตอบผิดไม่ได้ถูกตรวจสอบซ้ำด้วยคำตอบถูกและคำตอบผิดที่ได้จากข้อสอบข้ออื่น (Wright, 1992) เช่นเดียวกับแผนที่ภาวะสันนิษฐานที่จะไม่ถูกตรวจสอบความถูกต้องและเหมาะสมซ้ำด้วยข้อสอบข้ออื่น ดังนั้นจึงเกิดความไม่สอดคล้องกลมกลืนกันระหว่างข้อมูลคำตอบข้อสอบ มาตรฐานการเรียนรู้ และโครงสร้างความสามารถนักเรียน

ประการสุดท้ายคือ ข้อสอบเพียงข้อเดียวไม่สามารถให้ค่าประมาณความยากข้อสอบและความสามารถนักเรียนที่แม่นยำได้ เพราะค่าความคลาดเคลื่อนมาตรฐานจะสูงส่งผลให้ช่วงความเชื่อมั่นของค่าประมาณความสามารถกว้าง นั่นคือความแม่นยำของค่าประมาณความยากข้อสอบและความสามารถนักเรียนจะลดลง (Wright, 1992) และช่วงความเชื่อมั่นของค่าประมาณความสามารถจะครอบคลุมระดับความสามารถหลายระดับ ทำให้แผนที่ภาวะสันนิษฐานที่สร้างขึ้นจากข้อสอบเพียง

ข้อเดียว นอกจากจะไม่สามารถระบุได้ชัดเจนว่านักเรียนแต่ละคนนั้นมีความสามารถอยู่ในระดับใดแล้วยังไม่สามารถตีความผลการวัดได้อีกด้วย

แนวทางแก้ไขที่เป็นไปได้และเหมาะสมของโครงการวิจัยนี้คือ การยุบรวมแผนที่ภาวะสันนิษฐานรายมาตรฐานการเรียนรู้เป็นแผนที่ภาวะสันนิษฐานรายสาระการเรียนรู้ ซึ่งนอกจากจะทำให้มีข้อสอบจำนวนมากขึ้นต่อหนึ่งแผนที่ภาวะสันนิษฐานและแก้ปัญหาทั้งห้าประการข้างต้นแล้วยังคงรักษาความเป็นพหุมิติของแบบสอบ ONET ที่ประกอบด้วยเนื้อหาหลายส่วนและสอดคล้องกับรูปแบบรายงานผลคะแนนสอบ ONET ในปัจจุบันอีกด้วย

รายละเอียดของแผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์รายสาระการเรียนรู้และภาพรวมที่แสดงในภาพที่ 3.3 ถึง 3.11 มีแนวคิดสำคัญเหมือนกับโครงสร้างความยากข้อสอบและความสามารถที่นักเรียนใช้ทำข้อสอบ ONET ในภาพที่ 3.2 แต่ก็มีรายละเอียดบางอย่างไม่เหมือนกัน กล่าวคือ ประการแรก เมื่อนำข้อมูลตัวชี้วัดของมาตรฐานการเรียนรู้ซึ่งสัมพันธ์กับข้อสอบ ONET แต่ละข้อมาบรรยายความสามารถนักเรียนเพิ่มเติม ทำให้แผนที่ภาวะสันนิษฐานรายสาระการเรียนรู้ชัดเจนมากขึ้นว่าหลักการหรือความรู้ทางวิทยาศาสตร์ของแต่ละแผนที่ภาวะสันนิษฐานคืออะไร ประการที่สองคือ ระดับความสามารถจะมีชื่อเรียกแตกต่างกันและสอดคล้องกับระดับความยากข้อสอบด้วย ได้แก่ ขั้นสูง (Advanced) ขั้นเชี่ยวชาญ (Proficient) ขั้นพื้นฐาน (Basic) และขั้นต้น (Beginning) ซึ่งจะสัมพันธ์กับระดับความยากข้อสอบตั้งแต่ระดับยาก ค่อนข้างยาก ปานกลาง และง่าย ตามลำดับ ส่วนความสามารถขั้นต่ำกว่าขั้นต้น (Below Beginning) เป็นระดับที่กำหนดขึ้นในภายหลังสำหรับระบุตำแหน่งของนักเรียนที่ไม่สามารถตอบถูกข้อสอบง่ายได้ ประการที่สามคือ ลูกศรสองหัวชี้ขึ้นด้านบนและล่างใช้แทนสเกลต่อเนื่อง (continuous scale) ของปริมาณความสามารถที่เพิ่มขึ้นจากด้านล่างขึ้นด้านบนและปริมาณความสามารถที่ลดลงจากด้านบนลงด้านล่างด้วย นอกจากนี้ยังใช้แสดงขอบเขตความสามารถของนักเรียน โดยไม่ว่าจะมีนักเรียนคนใดก็ตามที่มีความสามารถเหนือกว่าขั้นสูงขึ้นไปก็จะพิจารณาให้นักเรียนกลุ่มดังกล่าวมีความสามารถในระดับเดียวกับขั้นสูง ในทำนองเดียวกัน ไม่ว่าจะมีนักเรียนคนใดที่มีความสามารถน้อยกว่าขั้นต่ำกว่าขั้นต้นก็จะพิจารณาให้นักเรียนกลุ่มนั้นมีความสามารถไม่แตกต่างจากระดับขั้นต่ำกว่าขั้นต้นเช่นกัน

สาระสำคัญของแผนที่ภาวะสันนิษฐานเชิงทฤษฎีในภาพที่ 3.3 – 3.11 นอกจากเป็นการบรรยายความสามารถนักเรียนแต่ละระดับว่ารู้เนื้อหาอะไรและมีทักษะการคิดเป็นอย่างไร ตำแหน่งของกลุ่มข้อสอบ ONET ยังแสดงตำแหน่งของนักเรียนที่มีโอกาสตอบข้อสอบแต่ละกลุ่มถูกอย่างน้อย 50% เช่น นักเรียนที่มีความสามารถขั้นพื้นฐานจะมีโอกาส 50% ที่จะตอบถูกข้อสอบข้อที่ 1, 2, และ 6 ของปีการศึกษา 2550 แต่จะมีโอกาสน้อยกว่า 50% ที่จะตอบถูกข้อ 4 และ 5 และมีโอกาสมากกว่า 50% ที่จะตอบถูกข้อ 3 และ 7 ของปีเดียวกัน

คำบรรยายความสามารถนักเรียน	ข้อสอบ		
	ปี 50	ปี 51	ปี 52
ขั้นสูง (Advanced)			
เข้าใจความสำคัญสารอาหารและกระบวนการเจริญเติบโตของต่อสิ่งมีชีวิต หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต การถ่ายทอดลักษณะทางพันธุกรรมและความหลากหลายของสิ่งมีชีวิต	50.4		52.3
เชื่อมโยงหลักการพื้นฐานและความรู้ทางวิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความข้อมูลทางวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับการเจริญเติบโตของสิ่งมีชีวิตและกระบวนการดำรงชีวิต แล้วสรุปเป็นข้อค้นพบที่ถูกต้อง			52.4
ขั้นเชี่ยวชาญ (Proficient)			
เข้าใจความสำคัญสารอาหารและกระบวนการเจริญเติบโตของต่อสิ่งมีชีวิต หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต การถ่ายทอดลักษณะทางพันธุกรรมและความหลากหลายของสิ่งมีชีวิต	50.5	51.1	52.6
เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ต่างๆ		51.3	52.27
ตีความข้อมูลทางวิทยาศาสตร์เกี่ยวกับการเจริญเติบโตของสิ่งมีชีวิตและกระบวนการดำรงชีวิต แล้วสรุปเป็นข้อค้นพบที่ถูกต้อง			52.37
ขั้นพื้นฐาน (Basic)			
มีความรู้พื้นฐานเกี่ยวกับสารอาหารและกระบวนการเจริญเติบโตของต่อสิ่งมีชีวิต หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต การถ่ายทอดลักษณะทางพันธุกรรมและความหลากหลายของสิ่งมีชีวิต ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ทั่วไป	50.1	51.2	52.25
ในบริบทต่างๆ หรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์เบื้องต้น	50.2	51.6	52.26
	50.6	51.7	
ขั้นต้น (Beginning)			
มีความรู้พื้นฐานเกี่ยวกับสารอาหารและกระบวนการเจริญเติบโตของต่อสิ่งมีชีวิต หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต การถ่ายทอดลักษณะทางพันธุกรรมและความหลากหลายของสิ่งมีชีวิต ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย	50.3	51.4	52.2
	50.7	51.5	52.5
ต่ำกว่าขั้นต้น (Below Beginning)			
มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับสารอาหารและกระบวนการเจริญเติบโตของต่อสิ่งมีชีวิต หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต การถ่ายทอดลักษณะทางพันธุกรรมและความหลากหลายของสิ่งมีชีวิต ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปแก้ปัญหาอย่างเหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย			

ภาพที่ 3.3 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

คำบรรยายความสามารถนักเรียน

ข้อสอบ

ปี 50 ปี 51 ปี 52

ขั้นสูง (Advanced)

เข้าใจความสัมพันธ์ระหว่างสิ่งมีชีวิตด้วยกันเองและสิ่งมีชีวิตกับสิ่งแวดล้อมในระบบนิเวศ และความสัมพันธ์แบบห่วงโซ่อาหารและสายใยอาหาร และเข้าใจสภาพแวดล้อมในท้องถิ่น และการรักษาทรัพยากรธรรมชาติและสิ่งแวดล้อม เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์ และความรู้ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความ ข้อมูลทางวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับความสัมพันธ์ของสิ่งต่างๆ ในระบบนิเวศแล้วสรุป เป็นข้อค้นพบที่ถูกต้อง

51.8

ขั้นเชี่ยวชาญ (Proficient)

เข้าใจความสัมพันธ์ระหว่างสิ่งมีชีวิตด้วยกันเองและสิ่งมีชีวิตกับสิ่งแวดล้อมในระบบนิเวศ และความสัมพันธ์แบบห่วงโซ่อาหารและสายใยอาหาร และเข้าใจสภาพแวดล้อมในท้องถิ่น และการรักษาทรัพยากรธรรมชาติและสิ่งแวดล้อม เชื่อมโยงหลักการพื้นฐานทางและนำไปใช้ แก้ปัญหาในสถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์เกี่ยวกับความสัมพันธ์ของสิ่ง ต่างๆ ในระบบนิเวศแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง

50.8 51.10 52.33

50.9

ขั้นพื้นฐาน (Basic)

มีความรู้พื้นฐานเกี่ยวกับความสัมพันธ์ระหว่างสิ่งมีชีวิตด้วยกันเองและสิ่งมีชีวิตกับสิ่งแวดล้อม ในระบบนิเวศ และความสัมพันธ์แบบห่วงโซ่อาหารและสายใยอาหาร รู้จักสภาพแวดล้อมใน ท้องถิ่นและวิธีรักษาทรัพยากรธรรมชาติและสิ่งแวดล้อม ใช้ความรู้พื้นฐานทางวิทยาศาสตร์ บางส่วนแก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์ เบื้องต้น

ขั้นต้น (Beginning)

มีความรู้พื้นฐานเกี่ยวกับความสัมพันธ์ระหว่างสิ่งมีชีวิตด้วยกันเองและสิ่งมีชีวิตกับสิ่งแวดล้อม ในระบบนิเวศ และความสัมพันธ์แบบห่วงโซ่อาหารและสายใยอาหาร รู้จักสภาพแวดล้อมใน ท้องถิ่นและวิธีดูแลรักษาทรัพยากรธรรมชาติและสิ่งแวดล้อม ใช้ความรู้พื้นฐานทาง วิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์ อย่างง่าย

51.9 52.1

ต่ำกว่าขั้นต้น (Below Beginning)

มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับความสัมพันธ์ ระหว่างสิ่งมีชีวิตด้วยกันเองและสิ่งมีชีวิตกับสิ่งแวดล้อมในระบบนิเวศ และความสัมพันธ์แบบ ห่วงโซ่อาหารและสายใยอาหาร และการรักษาทรัพยากรธรรมชาติและสิ่งแวดล้อม ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปใช้แก้ปัญหาได้อย่างเหมาะสม ไม่เข้าใจหรือ เข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย

ภาพที่ 3.4 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

คำบรรยายความสามารถนักเรียน		ข้อสอบ		
		ปี 50	ปี 51	ปี 52
↑	ขั้นสูง (Advanced)			
	เข้าใจสมบัติของสารและการเปลี่ยนแปลงสถานะของสาร ความสัมพันธ์ระหว่างสมบัติของ		51.14	
	สารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค การจำแนกสาร การเกิดสารละลายและ		51.16	
	การเกิดปฏิกิริยา และการเลือกใช้สารอย่างปลอดภัย เชื่อมโยงหลักการพื้นฐานทาง			
	วิทยาศาสตร์และความรู้ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน			
	ตีความข้อมูลทางวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับสมบัติของสารแล้วสรุปเป็นข้อค้นพบที่			
	ถูกต้อง			
	ขั้นเชี่ยวชาญ (Proficient)			
	เข้าใจสมบัติของสารและการเปลี่ยนแปลงสถานะของสาร ความสัมพันธ์ระหว่างสมบัติของ		51.18	
	สารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค การจำแนกสาร การเกิดสารละลายและ			
	การเกิดปฏิกิริยา และการเลือกใช้สารอย่างปลอดภัย เชื่อมโยงหลักการพื้นฐานทาง			
	วิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์			
	เกี่ยวกับสมบัติของสารแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง			
	ขั้นพื้นฐาน (Basic)	50.10	51.11	52.12
	มีความรู้พื้นฐานเกี่ยวกับสมบัติของสารและการเปลี่ยนแปลงสถานะของสาร ความสัมพันธ์	50.11	51.15	52.13
	ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค การจำแนกสาร การ	50.13	51.17	52.15
	เกิดสารละลายและการเกิดปฏิกิริยา และการเลือกใช้สารอย่างปลอดภัย ใช้ความรู้พื้นฐาน	50.15		52.16
	ทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือเกี่ยวกับข้อมูล	50.16		
	ทางวิทยาศาสตร์เบื้องต้น	50.18		
		50.19		
	ขั้นต้น (Beginning)			
	มีความรู้พื้นฐานเกี่ยวกับสมบัติของสารและการเปลี่ยนแปลงสถานะของสาร ความสัมพันธ์	50.12	51.12	52.8
	ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค การจำแนกสาร การ	50.14	51.13	52.9
	เกิดสารละลายและการเกิดปฏิกิริยา และการเลือกใช้สารอย่างปลอดภัย ใช้ความรู้พื้นฐาน	50.17		52.10
	ทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับข้อมูลทาง			52.11
	วิทยาศาสตร์อย่างง่าย			
	ต่ำกว่าขั้นต้น (Below Beginning)			
	มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับสมบัติของสาร			
	และการเปลี่ยนแปลงสถานะของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและ			
	แรงยึดเหนี่ยวระหว่างอนุภาค การจำแนกสาร การเกิดสารละลายและการเกิดปฏิกิริยา และ			
	การเลือกใช้สารอย่างปลอดภัย ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปใช้แก้ปัญหา			
	อย่างเหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย			
	↓			

ภาพที่ 3.5 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้สารและสมบัติของสาร

	ขั้นสูง (Advanced) เข้าใจธรรมชาติของความดันและแรงพุงของของเหลว การลอยตัวและการจมของวัตถุ การเคลื่อนที่ของวัตถุและแรงลัพธ์ และประโยชน์ของแรงเสียดทาน เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์และความรู้ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความข้อมูลวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับแรงและการเคลื่อนที่ของวัตถุในธรรมชาติแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง			
	ขั้นเชี่ยวชาญ (Proficient) เข้าใจธรรมชาติของความดันและแรงพุงของของเหลว การลอยตัวและการจมของวัตถุ การเคลื่อนที่ของวัตถุและแรงลัพธ์ และประโยชน์ของแรงเสียดทาน เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์เกี่ยวกับแรงและการเคลื่อนที่ของวัตถุในธรรมชาติแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง	52.17		52.34
	ขั้นพื้นฐาน (Basic) มีความรู้พื้นฐานเกี่ยวกับธรรมชาติของความดันและแรงพุงของของเหลว การลอยตัวและการจมของวัตถุ การเคลื่อนที่ของวัตถุและแรงลัพธ์ และประโยชน์ของแรงเสียดทาน ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์เบื้องต้น	50.20	51.20	52.7
	ขั้นต้น (Beginning) มีความรู้พื้นฐานเกี่ยวกับธรรมชาติของความดันและแรงพุงของของเหลว การลอยตัวและการจมของวัตถุ การเคลื่อนที่ของวัตถุและแรงลัพธ์ และประโยชน์ของแรงเสียดทาน ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย	50.22	51.19	52.18
	ต่ำกว่าขั้นต้น (Below Beginning) มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับธรรมชาติของความดันและแรงพุงของของเหลว การลอยตัวและการจมของวัตถุ การเคลื่อนที่ของวัตถุและแรงลัพธ์ และประโยชน์ของแรงเสียดทาน ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปใช้แก้ปัญหาอย่างเหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย			

ภาพที่ 3.6 แผนภาพสถานะสมรรถนะเชิงทฤษฎีของสาระการเรียนรู้แรงและการเคลื่อนที่

	ขั้นสูง (Advanced) เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การต่อเซลล์ไฟฟ้า การต่อหลอดไฟฟ้า และการต่อวงจรไฟฟ้าอย่างง่าย การเกิดเสียงและการเคลื่อนที่ของเสียง การสะท้อนและการหักเหของแสง การเปลี่ยนรูปพลังงานและผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์และความรู้ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความข้อมูลทางวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับพลังงานแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง			
	ขั้นเชี่ยวชาญ (Proficient) เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การต่อเซลล์ไฟฟ้า การต่อหลอดไฟฟ้า และการต่อวงจรไฟฟ้าอย่างง่าย การเกิดเสียงและการเคลื่อนที่ของเสียง การสะท้อนและการหักเหของแสง การเปลี่ยนรูปพลังงานและผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์เกี่ยวกับพลังงานแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง	50.25		52.35
		50.26		
		50.29		
	ขั้นพื้นฐาน (Basic) มีความรู้พื้นฐานเกี่ยวกับความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การต่อเซลล์ไฟฟ้า การต่อหลอดไฟฟ้า และการต่อวงจรไฟฟ้าอย่างง่าย การเกิดเสียงและการเคลื่อนที่ของเสียง การสะท้อนและการหักเหของแสง การเปลี่ยนรูปพลังงานและผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือเกี่ยวข้องกับข้อมูลทางวิทยาศาสตร์เบื้องต้น	50.23	51.24	52.20
		50.27	51.25	52.21
		50.28	51.27	52.22
	ขั้นต้น (Beginning) มีความรู้พื้นฐานเกี่ยวกับความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การต่อเซลล์ไฟฟ้า การต่อหลอดไฟฟ้า และการต่อวงจรไฟฟ้าอย่างง่าย การเกิดเสียงและการเคลื่อนที่ของเสียง การสะท้อนและการหักเหของแสง การเปลี่ยนรูปพลังงานและผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคย หรือเกี่ยวข้องกับข้อมูลทางวิทยาศาสตร์อย่างง่าย	50.24	51.22	52.19
			51.23	
			51.26	
	ต่ำกว่าขั้นต้น (Below Beginning) มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การต่อเซลล์ไฟฟ้า การต่อหลอดไฟฟ้า และการต่อวงจรไฟฟ้าอย่างง่าย การเกิดเสียงและการเคลื่อนที่ของเสียง การสะท้อนและการหักเหของแสง การเปลี่ยนรูปพลังงานและผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปแก้ปัญหาอย่างเหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย			

ภาพที่ 3.7 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้พลังงาน

	ขั้นสูง (Advanced) เข้าใจกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลกที่ส่งผลต่อภูมิประเทศ ภูมิอากาศ และสัณฐานของโลก ลักษณะและสมบัติของของหินและดิน การเกิดลมและสภาพภูมิอากาศ และผลกระทบของธรณีพิบัติภัยต่อมนุษย์และสภาพแวดล้อม เชื่อมโยงหลักการพื้นฐานทาง วิทยาศาสตร์และความรู้ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความข้อมูลทางวิทยาศาสตร์ที่ซับซ้อนเกี่ยวกับกระบวนการเปลี่ยนแปลงของโลกแล้วสรุปเป็น ข้อค้นพบที่ถูกต้อง		
	ขั้นเชี่ยวชาญ (Proficient) เข้าใจกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลกที่ส่งผลต่อภูมิประเทศ ภูมิอากาศ และสัณฐานของโลก ลักษณะและสมบัติของของหินและดิน การเกิดลมและสภาพภูมิอากาศ และผลกระทบของธรณีพิบัติภัยต่อมนุษย์และสภาพแวดล้อม เชื่อมโยงหลักการพื้นฐานทาง วิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาในสถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์ เกี่ยวกับกระบวนการเปลี่ยนแปลงของโลกแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง	50.31	51.32
		50.33	
	ขั้นพื้นฐาน (Basic) มีความรู้พื้นฐานเกี่ยวกับกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลกที่ส่งผลต่อ ภูมิประเทศ ภูมิอากาศ และสัณฐานของโลก ลักษณะและสมบัติของของหินและดิน การเกิด ลมและสภาพภูมิอากาศ และผลกระทบของธรณีพิบัติภัยต่อมนุษย์และสภาพแวดล้อม ใช้ ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือ เกี่ยวกับข้อมูลทางวิทยาศาสตร์เบื้องต้น	51.28	52.23
		51.29	
		51.30	
		51.31	
		51.33	
	ขั้นต้น (Beginning) มีความรู้พื้นฐานเกี่ยวกับกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลกที่ส่งผลต่อ ภูมิประเทศ ภูมิอากาศ และสัณฐานของโลก ลักษณะและสมบัติของของหินและดิน การเกิด ลมและสภาพภูมิอากาศ และผลกระทบของธรณีพิบัติภัยต่อมนุษย์และสภาพแวดล้อม ใช้ ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วนแก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับ ข้อมูลทางวิทยาศาสตร์อย่างง่าย	50.30	52.24
		50.32	52.30
		50.34	
		50.35	
	ต่ำกว่าขั้นต้น (Below Beginning) มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับกระบวนการ ต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลกที่ส่งผลต่อภูมิประเทศ ภูมิอากาศ และสัณฐานของ โลก ลักษณะและสมบัติของของหินและดิน การเกิดลมและสภาพภูมิอากาศ และผลกระทบ ของธรณีพิบัติภัยต่อมนุษย์และสภาพแวดล้อม ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ ไปแก้ปัญหาอย่างเหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์ อย่างง่าย		

ภาพที่ 3.8 แผนทึ่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

คำบรรยายความสามารถนักเรียน

ข้อสอบ

ปี 50 ปี 51 ปี 52

	ขั้นสูง (Advanced) เข้าใจลักษณะของระบบสุริยะ การเกิดทิดและปรากฏการณ์ขึ้นตกของดวงดาว การเกิดฤดู ข้างขึ้นข้างแรม สุริยุปราคาและจันทรุปราคา และความก้าวหน้าของเทคโนโลยีอวกาศและ ประโยชน์ต่อมนุษย์ เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์และความรู้ต่างๆ และนำไปใช้ แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน ตีความข้อมูลทางวิทยาศาสตร์ที่ซับซ้อน เกี่ยวกับดาราศาสตร์และอวกาศแล้วสรุปเป็นข้อค้นพบที่ถูกต้อง	50.40		
	ขั้นเชี่ยวชาญ (Proficient) เข้าใจลักษณะของระบบสุริยะ การเกิดทิดและปรากฏการณ์ขึ้นตกของดวงดาว การเกิดฤดู ข้างขึ้นข้างแรม สุริยุปราคาและจันทรุปราคา และความก้าวหน้าของเทคโนโลยีอวกาศและ ประโยชน์ต่อมนุษย์ เชื่อมโยงหลักการพื้นฐานทางวิทยาศาสตร์ต่างๆ และนำไปใช้แก้ปัญหาใน สถานการณ์ต่างๆ ตีความข้อมูลทางวิทยาศาสตร์เกี่ยวกับดาราศาสตร์และอวกาศแล้วสรุปเป็น ข้อค้นพบที่ถูกต้อง	50.38	51.35	52.36
	ขั้นพื้นฐาน (Basic) มีความรู้พื้นฐานเกี่ยวกับลักษณะของระบบสุริยะ การเกิดทิดและปรากฏการณ์ขึ้นตกของ ดวงดาว การเกิดฤดู ข้างขึ้นข้างแรม สุริยุปราคาและจันทรุปราคา และความก้าวหน้าของ เทคโนโลยีอวกาศและประโยชน์ต่อมนุษย์ ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วน แก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ หรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์เบื้องต้น	50.37	51.36	52.31
			51.37	
			51.39	
	ขั้นต้น (Beginning) มีความรู้พื้นฐานเกี่ยวกับลักษณะของระบบสุริยะ การเกิดทิดและปรากฏการณ์ขึ้นตกของ ดวงดาว การเกิดฤดู ข้างขึ้นข้างแรม สุริยุปราคาและจันทรุปราคา และความก้าวหน้าของ เทคโนโลยีอวกาศและประโยชน์ต่อมนุษย์ ใช้ความรู้พื้นฐานทางวิทยาศาสตร์บางส่วน แก้ปัญหาในสถานการณ์ที่คุ้นเคยหรือเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย	50.36	51.34	52.28
		50.39	51.38	52.29
			51.40	52.32
	ต่ำกว่าขั้นต้น (Below Beginning) มีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับลักษณะของ ระบบสุริยะ การเกิดทิดและปรากฏการณ์ขึ้นตกของดวงดาว การเกิดฤดู ข้างขึ้นข้างแรม สุริยุปราคาและจันทรุปราคา และความก้าวหน้าของเทคโนโลยีอวกาศและประโยชน์ต่อ มนุษย์ ไม่สามารถนำความรู้พื้นฐานทางวิทยาศาสตร์ไปแก้ปัญหอย่างเหมาะสม ไม่เข้าใจ หรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์อย่างง่าย			

ภาพที่ 3.9 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ดาราศาสตร์และอวกาศ

	ข้อสอบ		
	ปี 50	ปี 51	ปี 52
ขั้นสูง (Advanced)			
เข้าใจกระบวนการทางวิทยาศาสตร์และนำไปใช้สืบเสาะหาความรู้ รวมทั้งใช้ตรวจสอบผลที่เกิดขึ้น ใช้ข้อมูลหรือหลักฐานทางวิทยาศาสตร์ที่ถูกต้องและหลากหลายมาสนับสนุนหรือโต้แย้งข้อค้นพบ ใช้หลักการของกระบวนการทางวิทยาศาสตร์มาอธิบายหรือบรรยายปรากฏการณ์ทางธรรมชาติ เข้าใจความเกี่ยวข้องสัมพันธ์กันระหว่างกระบวนการทางวิทยาศาสตร์และความรู้ต่างๆ และนำกระบวนการดังกล่าวไปแก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน		51.14	52.3
			52.4
ขั้นเชี่ยวชาญ (Proficient)			
เข้าใจกระบวนการทางวิทยาศาสตร์และนำไปใช้สืบเสาะหาความรู้ รวมทั้งใช้ตรวจสอบผลที่เกิดขึ้น ใช้ข้อมูลหรือหลักฐานทางวิทยาศาสตร์ที่ถูกต้องมาสนับสนุนหรือโต้แย้งข้อค้นพบ ใช้หลักการของกระบวนการทางวิทยาศาสตร์มาอธิบายหรือบรรยายปรากฏการณ์ทางธรรมชาติ นำกระบวนการทางวิทยาศาสตร์ไปแก้ปัญหาในสถานการณ์ต่างๆ	50.8	51.3	52.27
	50.26	51.18	
	50.31		
	50.33		
ขั้นพื้นฐาน (Basic)			
มีความรู้พื้นฐานเกี่ยวกับกระบวนการทางวิทยาศาสตร์และนำไปใช้สืบเสาะหาความรู้ รวมทั้งใช้ตรวจสอบผลที่เกิดขึ้น ใช้ข้อมูลหรือหลักฐานทางวิทยาศาสตร์เบื้องต้นมาสนับสนุนหรือโต้แย้งข้อค้นพบ ใช้หลักการของกระบวนการทางวิทยาศาสตร์มาอธิบายหรือบรรยายปรากฏการณ์ทางธรรมชาติ ใช้กระบวนการทางวิทยาศาสตร์แก้ปัญหาในสถานการณ์ทั่วไปในบริบทต่างๆ	50.10	51.21	52.7
		51.24	52.13
		51.27	52.15
		51.28	
		51.30	
		51.31	
ขั้นต้น (Beginning)			
มีความรู้พื้นฐานเกี่ยวกับกระบวนการทางวิทยาศาสตร์และนำไปใช้สืบเสาะหาความรู้ รวมทั้งใช้ตรวจสอบผลที่เกิดขึ้น ใช้ข้อมูลหรือหลักฐานทางวิทยาศาสตร์อย่างง่ายมาสนับสนุนหรือโต้แย้งข้อค้นพบ ใช้หลักการของกระบวนการทางวิทยาศาสตร์มาอธิบายหรือบรรยายปรากฏการณ์ทางธรรมชาติ ใช้กระบวนการทางวิทยาศาสตร์แก้ปัญหาในสถานการณ์ที่คุ้นเคย			52.10
			52.11
ต่ำกว่าขั้นต้น (Below Beginning)			
มีความรู้พื้นฐานไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับกระบวนการทางวิทยาศาสตร์ นำกระบวนการทางวิทยาศาสตร์ไปใช้แก้ปัญหาไม่เหมาะสม ไม่เข้าใจหรือเข้าใจคลาดเคลื่อนเกี่ยวกับข้อมูลทางวิทยาศาสตร์ และไม่สามารถนำข้อมูลไปใช้อ้างอิงเพื่อสนับสนุนหรือโต้แย้งข้อค้นพบ			

ภาพที่ 3.10 แผนที่ภาวะสันนิษฐานเชิงทฤษฎีของสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

คำบรรยายความสามารถนักเรียน	ข้อสอบ						
ขั้นสูง (Advanced)							
เข้าใจหลักการและสาระสำคัญทางวิทยาศาสตร์ เชื่อมโยงความรู้ทางวิทยาศาสตร์กับความก้าวหน้าทางเทคโนโลยี เข้าใจความสัมพันธ์ระหว่างความรู้ทางวิทยาศาสตร์และกระบวนการทางวิทยาศาสตร์ ใช้ความรู้และกระบวนการทางวิทยาศาสตร์แก้ปัญหาในสถานการณ์ที่หลากหลายและซับซ้อน เข้าใจข้อมูลทางวิทยาศาสตร์และนำไปใช้เป็นหลักฐานอ้างอิงเพื่อสนับสนุนหรือโต้แย้งข้อค้นพบ	50.4	50.40	51.8	51.14	51.16	52.3	
	52.4						
ขั้นเชี่ยวชาญ (Proficient)							
เข้าใจหลักการและสาระสำคัญทางวิทยาศาสตร์ ความสัมพันธ์ระหว่างความรู้ทางวิทยาศาสตร์และเทคโนโลยีกับกระบวนการทางวิทยาศาสตร์ ใช้ความรู้และกระบวนการทางวิทยาศาสตร์เพื่อแก้ปัญหาในสถานการณ์ที่หลากหลาย เข้าใจข้อมูลทางวิทยาศาสตร์และนำไปใช้เป็นหลักฐานอ้างอิงเพื่อสนับสนุนหรือโต้แย้งข้อค้นพบ	50.5	50.8	50.9	50.25	50.26	50.29	
	50.31	50.33	50.38	51.1	51.3	51.10	
	51.18	51.32	51.35	52.6	52.7	52.17	
	52.27	52.33	52.34	52.35	52.36	52.37	
ขั้นพื้นฐาน (Basic)							
มีความรู้พื้นฐานทางวิทยาศาสตร์และเทคโนโลยี ใช้กระบวนการทางวิทยาศาสตร์สืบเสาะหาความรู้และตรวจสอบข้อค้นพบ ใช้ความรู้และกระบวนการทางวิทยาศาสตร์เพื่อแก้ปัญหาในสถานการณ์ทั่วไป ในบริบทต่างๆ และนำข้อมูลหรือหลักฐานทางวิทยาศาสตร์เบื้องต้นมาอ้างอิงเพื่อสนับสนุนหรือโต้แย้งข้อค้นพบ	50.1	50.2	50.6	50.10	50.11	50.13	
	50.15	50.16	50.18	50.19	50.20	50.21	
	50.23	50.27	50.28	50.37	51.2	51.6	
	51.7	51.11	51.15	51.17	51.20	51.21	
	51.24	51.25	51.27	51.28	51.29	51.30	
	51.31	51.33	51.36	51.37	51.39	52.7	
	52.12	52.13	52.14	52.15	52.16	52.20	
	52.21	52.22	52.23	52.25	52.26	52.31	
ขั้นต้น (Beginning)							
มีความรู้พื้นฐานทางวิทยาศาสตร์และเทคโนโลยี ใช้กระบวนการทางวิทยาศาสตร์สืบเสาะหาความรู้และตรวจสอบข้อค้นพบ ใช้ความรู้และกระบวนการทางวิทยาศาสตร์เพื่อแก้ปัญหาในสถานการณ์ที่คุ้นเคย	50.3	50.7	50.12	50.14	50.17	50.22	
	50.24	50.30	50.32	50.34	50.35	50.36	
	50.39	51.4	51.5	51.9	51.12	51.13	
	51.19	51.22	51.23	51.26	51.34	51.38	
	51.40	52.1	52.2	52.5	52.8	52.9	52.10
	52.11	52.18	52.19	52.24	52.28	52.29	
	52.30	52.32					
ต่ำกว่าขั้นต้น (Below Beginning)							
มีความรู้พื้นฐานไม่เพียงพอหรือเข้าใจคลาดเคลื่อนเกี่ยวกับหลักการและสาระสำคัญทางวิทยาศาสตร์ ไม่สามารถนำข้อมูลไปใช้อ้างอิงเพื่อสนับสนุนหรือโต้แย้งข้อค้นพบ นำกระบวนการทางวิทยาศาสตร์ไปใช้แก้ปัญหาไม่เหมาะสม							

ภาพที่ 3.11 แผนทีภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์

โมเดลการวัดเอกมิตีและพหุมิตี

จากที่ได้กล่าวไว้ในตอนต้น โมเดลการวัดที่พัฒนาขึ้นในโครงการวิจัยนี้แบ่งออกเป็นสองกลุ่ม ได้แก่ โมเดลเอกมิตีและโมเดลพหุมิตี โดยโมเดลการวัดแต่ละกลุ่มจะมีรูปแบบการให้คะแนนคำตอบแบบสองค่าและแบบหลายค่า ตารางที่ 3.6 แสดงให้เห็นว่าเนื้อหาของบทที่ 3 ซึ่งเกี่ยวข้องกับการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานเชิงทฤษฎีจะใช้วิธีวิเคราะห์แบบเอกมิตี ส่วนการตรวจสอบความสอดคล้องกลมกลืนของข้อมูลคำตอบข้อสอบ ONET กับโมเดลการวัดในบทที่ 4 จะใช้ทั้งวิธีวิเคราะห์แบบเอกมิตีและพหุมิตี

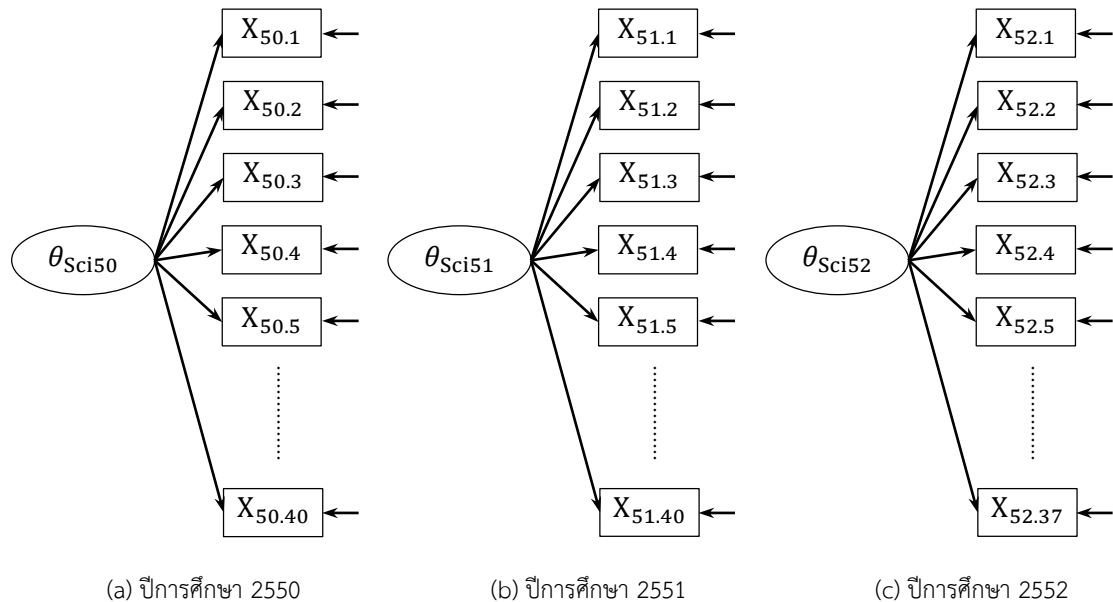
ตารางที่ 3.6 โมเดลการวัดความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6

วิธีวิเคราะห์	การตรวจสอบคุณภาพของแผนที่ภาวะสันนิษฐาน	การตรวจสอบความสอดคล้องกลมกลืนของข้อมูลกับโมเดล
วิธีวิเคราะห์แบบเอกมิตี		
รวมมิตีย่อย	โมเดล UniCM	โมเดล UniFit
แยกมิตีย่อย	โมเดล ConsUniCM	-
วิธีวิเคราะห์แบบพหุมิตี		
พหุมิตีระหว่างข้อสอบ	-	โมเดล BimFit
พหุมิตีภายในข้อสอบ	-	โมเดล WimFit

เมื่อพิจารณากลุ่มโมเดลประเภทต่างๆ ในตารางที่ 3.6 พบว่าโมเดลการวัดอย่างง่ายที่สุดคือ โมเดลเอกมิตีที่กำหนดให้ข้อสอบทุกข้อของแต่ละปีการศึกษาวัดความสามารถทางวิทยาศาสตร์เพียงมิติเดียว โมเดลนี้จึงเป็นโมเดลฐาน (baseline model) สำหรับเปรียบเทียบความสอดคล้องกลมกลืนกับโมเดลแข่งขัน (competing model) ภาพที่ 3.12 บรรยายข้อกำหนดลักษณะโมเดล (model specification) ของโมเดล UniCM และโมเดล UniFit ที่ความสามารถนักเรียนแต่ละคนของแต่ละปีการศึกษาไม่ถูกแยกเป็นความสามารถรายสาระการเรียนรู้ ความแตกต่างระหว่างสองโมเดลนี้คือ โมเดล UniCM จะถูกใช้ตรวจสอบคุณภาพของแผนที่ภาวะสันนิษฐานเชิงทฤษฎีตามภาพที่ 3.11 ด้วยกลุ่มตัวอย่างขนาด 5% จากฐานข้อมูลวิเคราะห์ ส่วนโมเดล UniFit จะใช้ตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลคำตอบข้อสอบของกลุ่มตัวอย่างขนาด 10% จากฐานข้อมูลวิเคราะห์เพื่อเป็นเกณฑ์เปรียบเทียบกับโมเดลการวัดอื่นๆ

ความแตกต่างระหว่างโมเดล a, b และ c ในภาพที่ 3.12 นอกจากแสดงให้เห็นว่าแต่ละโมเดลใช้ข้อสอบ ONET ไม่เหมือนกันแล้ว ยังแสดงถึงจำนวนข้อสอบและรูปแบบการให้คะแนนคำตอบที่ไม่เหมือนกันด้วยโดยเฉพาะข้อสอบข้อที่ 33-37 ในปีการศึกษา 2552 ที่กำหนดให้นักเรียนเลือกคำตอบสองตัวเลือกซึ่งรูปแบบการให้คะแนนดั้งเดิมนั้นจะไม่ให้คะแนนกับคำตอบถูกบางส่วน แต่

สำหรับโครงการวิจัยนี้จะให้คะแนนแบบหลายค่ากับคำตอบถูกบางส่วนตามความเห็นของผู้เชี่ยวชาญ โดยคะแนนบางส่วนนี้จะเท่ากับคะแนนในระดับที่ต่ำกว่าหนึ่งระดับของคะแนนที่ควรจะได้ของข้อนั้น ดังนั้นโมเดล c จึงประกอบด้วยสองโมเดลย่อยคือ โมเดลสำหรับข้อมูลคะแนนคำตอบแบบสองค่า และโมเดลสำหรับข้อมูลคะแนนคำตอบแบบหลายค่า



ภาพที่ 3.12 โมเดลเอกมิตีแบบรวมมิติความสามารถทางวิทยาศาสตร์ของนักเรียน

หมายเหตุ 1) θ_{Sci50} , θ_{Sci51} และ θ_{Sci52} หมายถึงค่าประมาณความสามารถทางวิทยาศาสตร์ของนักเรียนปีการศึกษา 2550, 2551 และ 2552 ตามลำดับ

2) $X_{50.1}$ ไปจนถึง $X_{52.37}$ หมายถึง คะแนนหรือคำตอบที่สังเกตได้ของข้อสอบแต่ละข้อตั้งแต่ข้อตั้งข้อสอบ ONET ข้อที่ 1 ปีการศึกษา 2550 ไปจนถึงข้อที่ 37 ปีการศึกษา 2552

ภาพที่ 3.12 แสดงโมเดลเอกมิตีความสามารถนักเรียนปีการศึกษา 2550 ถึง 2552 ค่าความยากข้อสอบ δ_i และความสามารถนักเรียน θ_n จะถูกตรวจสอบคุณสมบัติเชิงโครงสร้างด้วยโปรแกรม ConQuest 3.0 (Adams, Wu, & Wilson, 2012) ข้อมูลคำตอบข้อสอบ ONET จะถูกนำมาสร้างเป็นสเกลด้วยโมเดลของราส์ช (Rasch, 1960) และประมาณค่าพารามิเตอร์ต่างๆ ด้วยเทคนิคการประมาณค่าภาวะน่าจะเป็นสูงสุดชายขอบ (marginal maximum likelihood) ในกรณีการวัดแบบเอกมิตี คะแนนรวมจากข้อสอบทุกข้อของแบบสอบแต่ละปีจะเป็นค่าสถิติเพียงพอ (sufficient statistics: Fischer, 1995; Rasch, 1960) สำหรับประมาณความสามารถทางวิทยาศาสตร์ของนักเรียนเพียงค่าเดียว ฟังก์ชันความสัมพันธ์ระหว่างพารามิเตอร์ความยากข้อสอบและคะแนนคำตอบแต่ละข้อจะอยู่ในรูปของสมการโมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์เชิงสุ่มจากการวัดแบบมิตีเดียว (Random Coefficients Multinomial Logit) หรือโมเดล RCML (Wilson & Adams,

1995; Wilson & Wang, 1995) และโปรแกรม ConQuest จะประมาณค่าพารามิเตอร์ความยากข้อสอบ ค่าเฉลี่ยประชากรและค่าความแปรปรวนประชากรสำหรับโมเดลเอกมิตินั้น ข้อตกลงเบื้องต้นที่สำคัญของโมเดล RCML คือผลรวมของค่าพารามิเตอร์ความยากข้อสอบจะถูกจำกัดให้เท่ากับศูนย์ ซึ่งเท่ากับว่าค่าเฉลี่ยของความสามารถนักเรียนในประชากรจะมีค่าเท่ากับศูนย์ด้วย (Briggs & Wilson, 2003)

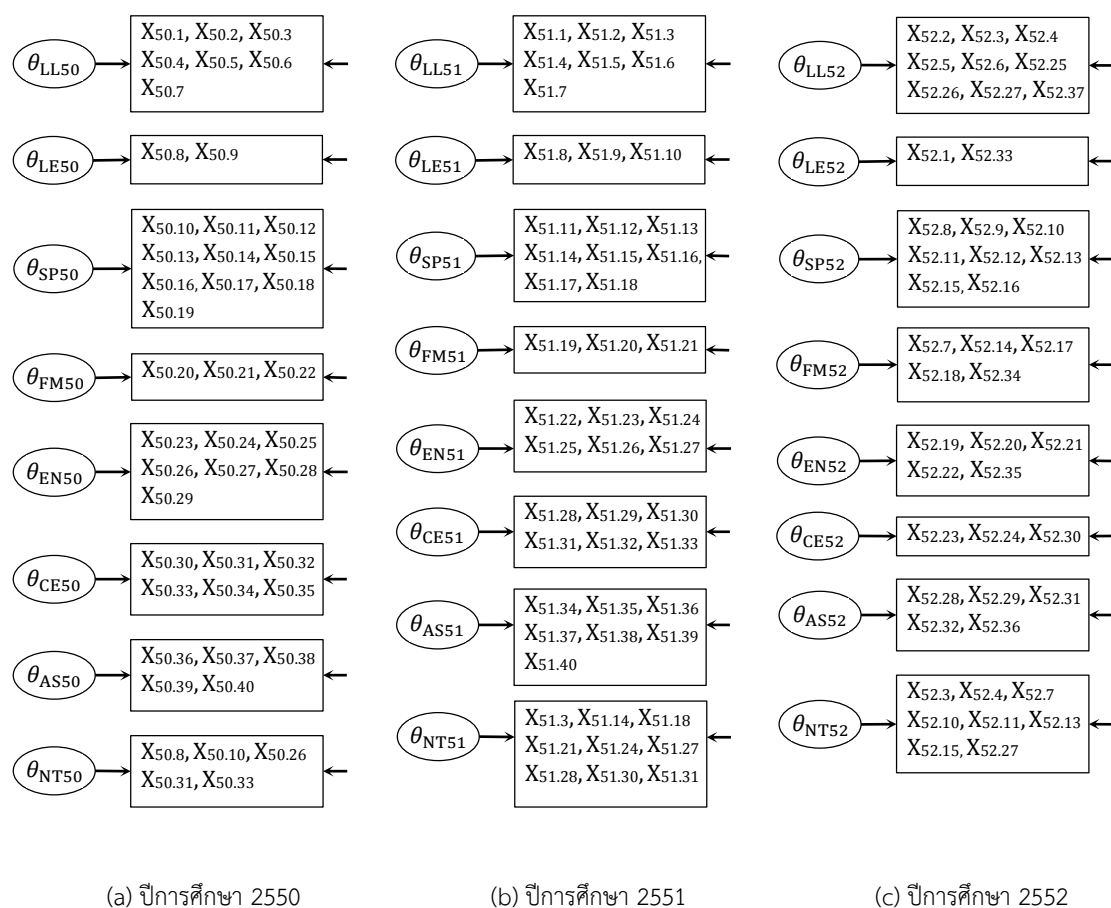
โมเดลในภาพที่ 3.12 เป็นโมเดลที่ใช้ประมาณค่าพารามิเตอร์ต่างๆ จากคำตอบข้อสอบของแต่ละปีการศึกษา ตัวอย่างเช่น สัญลักษณ์ θ_{Sci} หมายถึงค่าประมาณความสามารถทางวิทยาศาสตร์โดยรวมของนักเรียน n ที่ทำข้อสอบในแต่ละปีการศึกษา ดังนั้น θ_{Sci50} , θ_{Sci51} , และ θ_{Sci52} หมายถึงค่าประมาณความสามารถของนักเรียนแต่ละคนในแต่ละปีการศึกษา นอกจากนี้แต่ละโมเดลยังแสดงอิทธิพลทางตรงของ θ_{Sci} ต่อคำตอบข้อสอบที่สังเกตได้ (ลูกศรที่ชี้จาก θ_{Sci} ไปยัง X) ซึ่งเป็นผลรวมจากคะแนนคำตอบแต่ละข้อ ดังนั้น θ_{Sci50} จึงเป็นอิทธิพลทางตรงต่อคะแนนรวมที่ได้จากคำตอบข้อที่ 1 ($X_{50.1}$) ถึงข้อที่ 40 ($X_{50.40}$) θ_{Sci51} เป็นอิทธิพลทางตรงต่อคะแนนรวมที่ได้จากคำตอบข้อที่ 1 ($X_{51.1}$) ถึงข้อที่ 40 ($X_{51.40}$) และ θ_{Sci52} เป็นอิทธิพลทางตรงต่อคะแนนรวมที่ได้จากคำตอบข้อที่ 1 ($X_{52.1}$) ถึงข้อที่ 37 ($X_{52.37}$)

เมื่อพิจารณาจำนวนพารามิเตอร์ที่จะประมาณได้จากโมเดล RCML ที่มีวิธีให้คะแนนแตกต่างกันพบว่า โมเดลที่ให้คะแนนคำตอบแบบหลายค่าจะมีจำนวนพารามิเตอร์มากกว่าแบบสองค่า โดยจะมีค่าพารามิเตอร์ความยากลำดับขั้นทำหน้าที่เป็นเส้นแบ่งความสามารถนักเรียนเป็นระดับต่างๆ ข้อดีของโมเดลเอกมิตินี้อาศัยหลายประการ กล่าวคือ ความสามารถนักเรียนจะถูกสรุปรวมเป็นตัวเลขเดียวจึงตีความผลการวัดง่ายและนอกจากนี้ค่าความเที่ยงแบบ EAP/PV ของผลการวัดยังมีค่าสูงด้วย (Briggs & Wilson, 2003) เนื่องจากเมื่อมีข้อสอบที่วัดตัวแปรแฝงนั้นมากขึ้น คำตอบข้อสอบแต่ละข้อจะถูกตรวจสอบซ้ำด้วยข้อสอบข้ออื่นที่วัดตัวแปรแฝงเดียวกัน (Wright, 1992) แต่ในขณะเดียวกันโมเดลเอกมิตินี้จะสูญเสียสารสนเทศความสามารถนักเรียนรายด้านซึ่งสัมพันธ์กับมิติต่างๆ ที่ข้อสอบแต่ละข้ออาจจะวัดเนื้อหาไม่เหมือนกันก็ได้ (Briggs & Wilson, 2003)

โมเดลเอกมิตินี้มีอีกประเภทหนึ่งที่แสดงในตารางที่ 3.6 คือโมเดล ConsUniCM หรือโมเดลเอกมิติแบบแยกราชมิติ (consecutive unidimensional model: Briggs & Wilson, 2003) โมเดลนี้จะใช้ตรวจสอบคุณภาพของแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่แสดงในภาพที่ 3.3 ถึง 3.10 ลักษณะของโมเดลนี้คือจะใช้คะแนนรวมที่สังเกตได้จากข้อสอบที่วัดแต่ละสาระการเรียนรู้ประมาณค่าความสามารถนักเรียน ข้อกำหนดลักษณะโมเดลแบบแยกมิติของแต่ละปีการศึกษาแสดงในภาพที่ 3.13a, 3.13b และ 3.13c โดยที่โมเดล 3.13c จะถูกตรวจสอบด้วยข้อมูลคำตอบที่ให้คะแนนแบบสองค่าและแบบหลายค่าเช่นเดียวกับโมเดล 3.12c

โมเดล a, b และ c ในภาพที่ 3.13 จะถูกใช้ประมาณค่าความสามารถนักเรียนและค่าพารามิเตอร์ความยากข้อสอบเช่นเดียวกับโมเดลในภาพที่ 3.12 แต่จะแยกวิเคราะห์ข้าราชการการเรียนรู้ด้วยโมเดล RCML ข้อดีของวิธีวิเคราะห์นี้จะทำให้ได้สเกลความสามารถนักเรียนรายสาระ

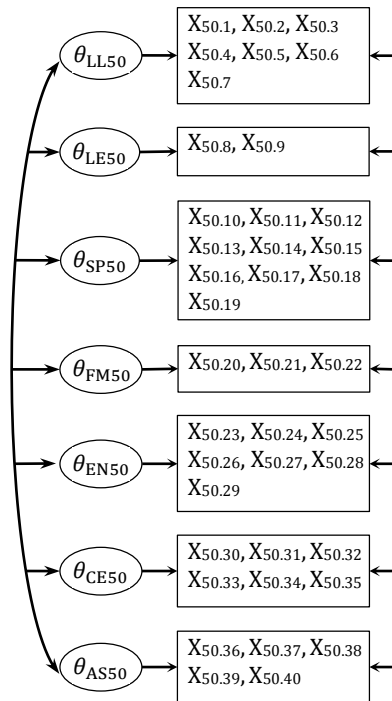
การเรียนรู้และค่าความคลาดเคลื่อนมาตรฐานของแต่ละสาระการเรียนรู้ด้วย นั่นคือผู้วิจัยจะได้สารสนเทศของนักเรียนและข้อสอบที่ชัดเจนและละเอียดมากขึ้น แต่อย่างไรก็ตามวิธีวิเคราะห์ดังกล่าวจำเป็นต้องมีข้อสอบจำนวนมากที่วัดแต่ละสาระการเรียนรู้ นั่น เพราะถ้ามีข้อสอบน้อยที่วัดแต่ละมิติน้อย ค่าความคลาดเคลื่อนมาตรฐานจะมากกว่าโมเดลเอกมิตินี้ที่รวมข้อสอบทุกข้อให้วัดเพียงมิติเดียว (Briggs & Wilson, 2003)



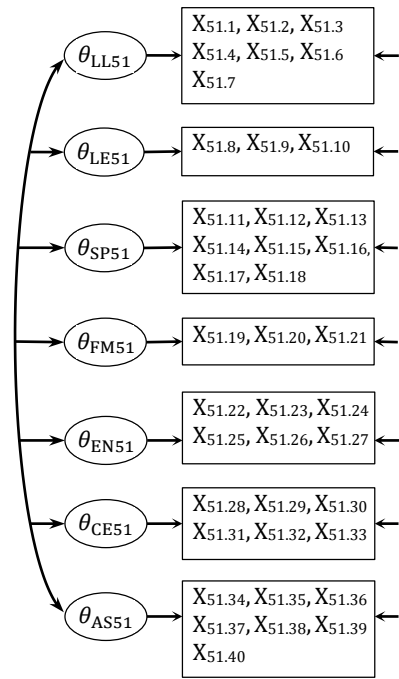
ภาพที่ 3.13 โมเดลเอกมิตินี้แบบแยกรายมิติความสามารถทางวิทยาศาสตร์ของนักเรียน

หมายเหตุ 1) เพื่อกระชับรายละเอียดข้อกำหนดคุณลักษณะโมเดล ผู้วิจัยจึงยุบรวมกรอบสี่เหลี่ยมย่อยที่แสดงค่าสังเกตได้ของข้อสอบแต่ละข้อที่วัดสาระการเรียนรู้เดียวกันเป็นกรอบสี่เหลี่ยมใหญ่เพียงอันเดียว และใช้เครื่องหมาย , คั่นระหว่างค่าสังเกตได้เพื่อแสดงว่าค่าสังเกตได้เหล่านั้นเป็นอิสระจากกัน

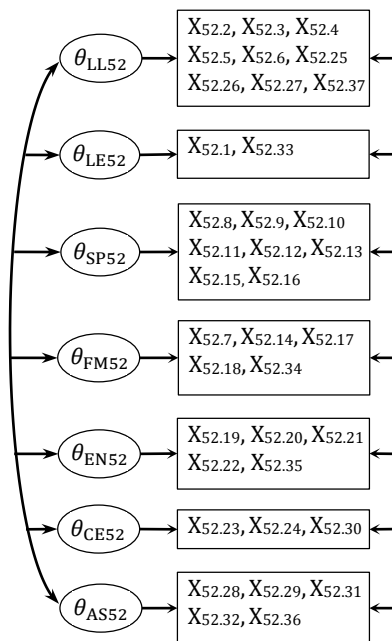
2) ค่าประมาณความสามารถนักเรียนรายสาระการเรียนรู้ ได้แก่ สิ่งมีชีวิตกับกระบวนการดำรงชีวิต (Living things and life processes: θ_{LL}) ชีวิตกับสิ่งแวดล้อม (Life and the Environment: θ_{LE}) สารและสมบัติของสาร (Substances and properties of substances: θ_{SP}) แรงและการเคลื่อนที่ (Forces and motion: θ_{FM}) พลังงาน (Energy: θ_{EN}) กระบวนการเปลี่ยนแปลงของโลก (Change processes of the Earth: θ_{CE}) ดาราศาสตร์และอวกาศ (Astronomy and space: θ_{AS}) และธรรมชาติของวิทยาศาสตร์และเทคโนโลยี (Nature of science and technology: θ_{NT})



(a) ปีการศึกษา 2550



(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552

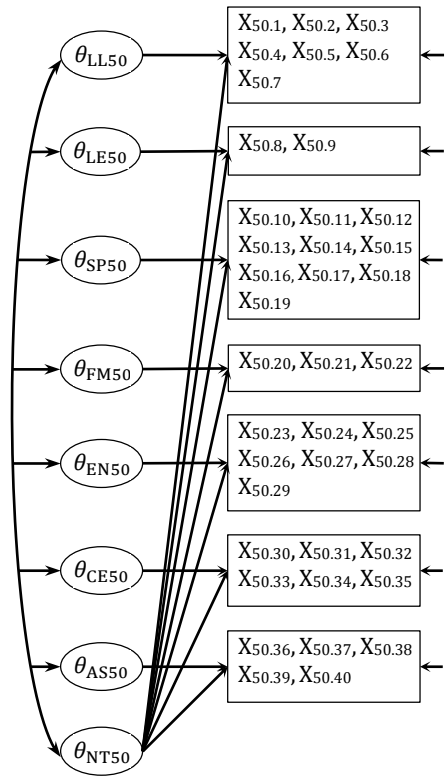
ภาพที่ 3.14 โมเดลพหุมิติระหว่างข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียน

หมายเหตุ เส้นโค้งที่ประกอบด้วยหัวลูกศรย่อยๆ ที่อยู่ทางด้านซ้ายของแต่ละโมเดลแสดงถึงความสัมพันธ์กันของค่าประมาณความสามารถนักเรียนของแต่ละสาระการเรียนรู้

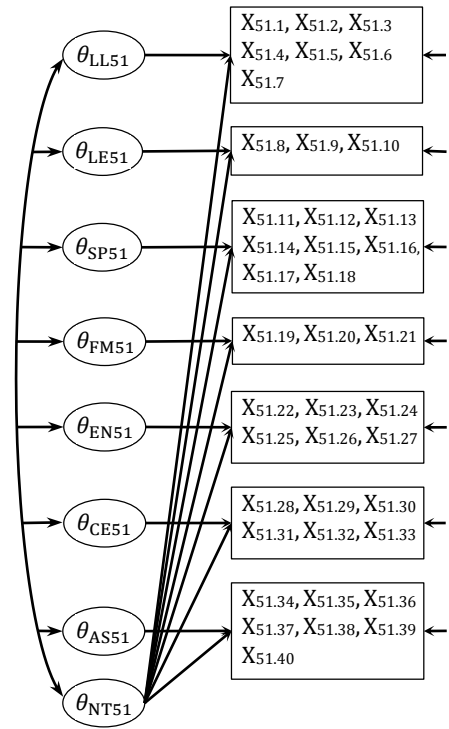
โมเดลกลุ่มที่สองคือ โมเดลพหุมิติระหว่างข้อสอบและโมเดลพหุมิติภายในข้อสอบ ลักษณะสำคัญของโมเดลพหุมิติที่สร้างขึ้นภายในกรอบโมเดล MRCML (Adams, Wilson, & Wang, 1997) คือเป็นการผสมผสานโมเดลเอกมิติแบบรายมิตีย่อยให้เป็นโมเดลเดียวกันโดยให้ค่าประมาณความสามารถนักเรียนรายสาระการเรียนรู้สัมพันธ์กัน นั่นหมายความว่าสารสนเทศต่างๆ ที่เกิดขึ้นจากแต่ละมิตีย่อยจะไม่สูญหายไปในการบวนการประมาณค่าพารามิเตอร์ต่างๆ

ภาพที่ 3.14 แสดงโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit ในตารางที่ 3.6 โมเดลนี้จะใช้ประเมินความสอดคล้องกลมกลืนกับข้อมูลคำตอบข้อสอบ ONET และนำผลความสอดคล้องกลมกลืนที่ได้ไปเปรียบเทียบกับโมเดล UniFit ลักษณะสำคัญของโมเดล BimFit คือข้อสอบแต่ละข้อจะวัดเพียงสาระการเรียนรู้เดียว นั่นคือความสามารถนักเรียนจะถูกแบ่งออกเป็น 7 มิตีย่อยที่สัมพันธ์กันโดยสันนิษฐานให้สาระการเรียนรู้ที่ 8 เกี่ยวกับธรรมชาติของวิทยาศาสตร์และเทคโนโลยีแทรกอยู่ในสาระการเรียนรู้ที่ 1 ถึง 7 ข้อสันนิษฐานนี้เกิดขึ้นเนื่องจากรายละเอียดของมาตรฐานการเรียนรู้ของแต่ละสาระการเรียนรู้ประกอบด้วยเนื้อหาสองส่วนคือ ความรู้ทางวิทยาศาสตร์และทักษะที่จำเป็น (ดูรายละเอียดในตาราง A1 ของภาคผนวก) โดยทักษะที่จำเป็นเหล่านี้สมมติให้เหมือนกับทักษะกระบวนการทางวิทยาศาสตร์ของสาระการเรียนรู้ที่ 8 นอกจากนี้ยังมีข้อสังเกตเกี่ยวกับข้อกำหนดลักษณะโมเดล BimFit คือ ข้อสอบแต่ละข้อไม่ได้วัดเนื้อหาทั้งสองส่วนของแต่ละสาระการเรียนรู้ในทางตรงกันข้ามชุดข้อสอบที่วัดสาระการเรียนรู้ต่างๆ จะร่วมกันวัดทั้งความรู้และทักษะที่จำเป็น การตีความผลการวัดที่เกิดขึ้นจากโมเดล BimFit จึงบรรยายว่านักเรียนมีความสามารถทางวิทยาศาสตร์ในแต่ละสาระการเรียนรู้เป็นอย่างไรทั้งในด้านเนื้อหาความรู้และทักษะไปพร้อมๆ กัน และผลวิเคราะห์ที่ได้จากโมเดลนี้จะไม่มีค่าประมาณความสามารถนักเรียนตามสาระการเรียนรู้ที่ 8 จึงทำให้ขาดความสมบูรณ์เมื่อสรุปอ้างอิงความสามารถทางวิทยาศาสตร์ของนักเรียนตามที่กำหนดในหลักสูตรแกนกลางฯ พ.ศ. 2551

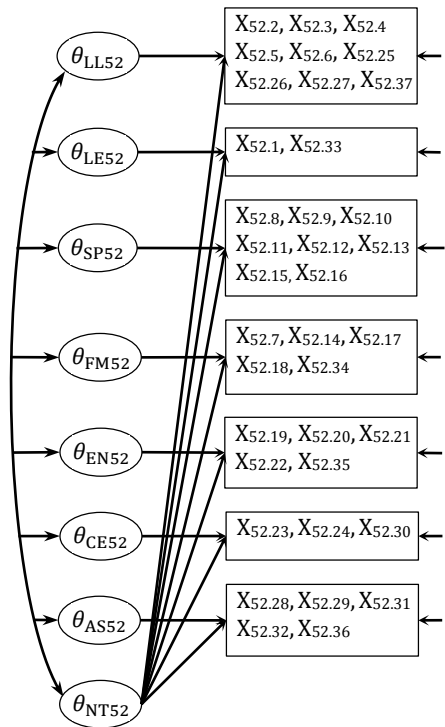
ภาพที่ 3.15 และ 3.16 แสดงโมเดลพหุมิติภายในข้อสอบหรือโมเดล WimFit ในตารางที่ 3.6 โมเดลนี้จะถูกประเมินความสอดคล้องกลมกลืนกับข้อมูลคำตอบข้อสอบและเปรียบเทียบความสอดคล้องกลมกลืนที่ได้กับโมเดล UniFit และโมเดล BimFit ลักษณะสำคัญของโมเดล WimFit คือจะมีข้อสอบบางข้อที่วัดมากกว่าหนึ่งสาระการเรียนรู้ เมื่อพิจารณาโมเดลในภาพที่ 3.15 จะพบว่าข้อสอบทุกข้อถูกกำหนดว่าวัดสาระการเรียนรู้ที่ 8 ด้วยแทนที่จะกำหนดให้สาระการเรียนรู้ที่ 8 แทรกอยู่ในสาระการเรียนรู้ที่ 1-7 การออกแบบโมเดลการวัดแบบนี้จึงตรงกันข้ามกับโมเดล BimFit และความสามารถทางวิทยาศาสตร์ของนักเรียนจะถูกแบ่งออกเป็น 8 มิตีย่อย ทำให้ผู้วิจัยได้สารสนเทศเพิ่มเติมเกี่ยวกับความสามารถนักเรียนและผลการวัดจะนำไปสู่ข้อสรุปที่ว่านักเรียนมีความรู้วิทยาศาสตร์และทักษะกระบวนการทางวิทยาศาสตร์มากขึ้นเพียงใด ข้อสังเกตที่น่าสนใจเกี่ยวกับโมเดลในภาพที่ 3.15 คือ ผู้วิจัยขาดหลักฐานจากผู้สร้างข้อสอบ ONET เพื่อยืนยันว่าในการบวนการสร้างข้อสอบนั้นข้อสอบ ONET แต่ละข้อถูกออกแบบให้วัดสาระการเรียนรู้ที่ 8 อย่างไร และอาจจะเป็นไปได้สูงที่จะมีข้อสอบบางข้อไม่วัดสาระการเรียนรู้ที่ 8 โมเดลในภาพที่ 3.15 จึงอาจจะเกิดปัญหา



(a) ปีการศึกษา 2550

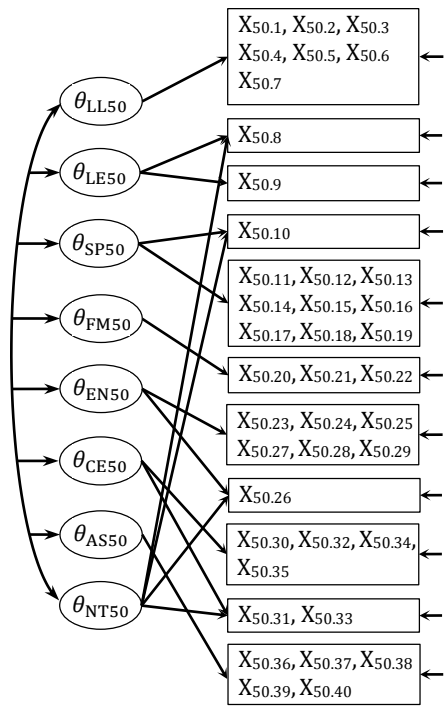


(b) ปีการศึกษา 2551

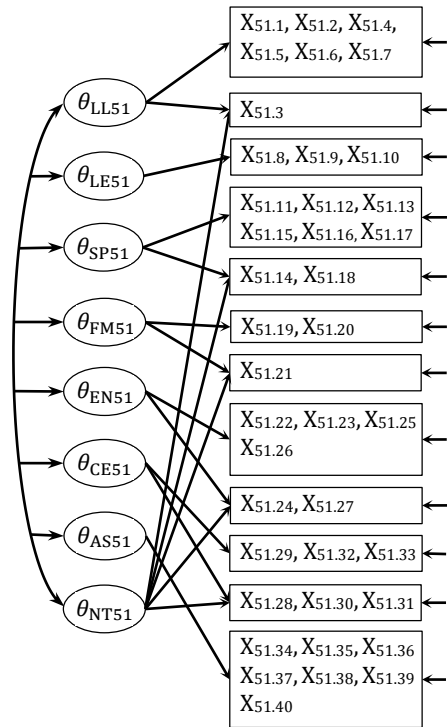


(c) ปีการศึกษา 2552

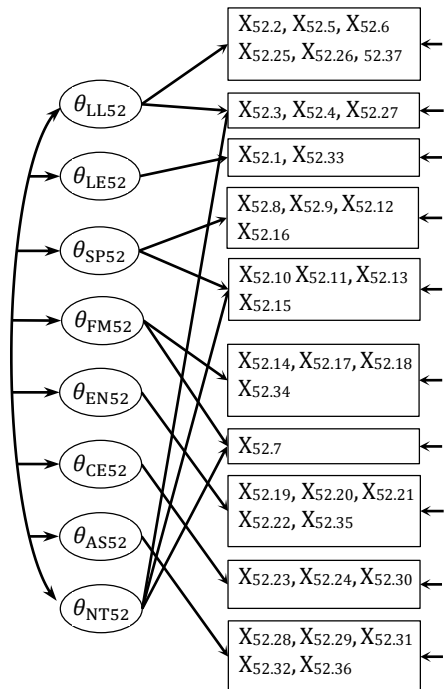
ภาพที่ 3.15 โมเดลพหุมิติภายในข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียนที่กำหนดให้
ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8



(a) ปีการศึกษา 2550



(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552

ภาพที่ 3.16 โมเดลพหุมิติภายในข้อสอบความสามารถทางวิทยาศาสตร์ของนักเรียนที่กำหนดให้
ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8

การกำหนดลักษณะโมเดลคลาดเคลื่อน (model misspecification) เมื่อพิจารณาโมเดลในภาพที่ 3.16 จะพบว่ามีข้อสอบบางข้อถูกกำหนดให้วัดสาระการเรียนรู้ที่ 8 ตามผลประเมินโดยผู้เชี่ยวชาญของโครงการวิจัย (ดูรายละเอียดในตารางที่ A2 ในภาคผนวก) และความสามารถทางวิทยาศาสตร์ของนักเรียนจะถูกแบ่งออกเป็น 8 มิติ ทำให้ผู้วิจัยได้สารสนเทศครบถ้วนเกี่ยวกับความสามารถนักเรียนตามหลักสูตรแกนกลางฯ พ.ศ. 2551 ทั้งในด้านความรู้วิทยาศาสตร์และทักษะกระบวนการทางวิทยาศาสตร์ ความแตกต่างระหว่างโมเดลในภาพที่ 3.16 และ 3.15 คือ ผู้วิจัยมีหลักฐานจากผู้เชี่ยวชาญมาสนับสนุนว่าข้อสอบข้อใดวัดหรือไม่วัดสาระการเรียนรู้ที่ 8 และอาจจะช่วยลดปัญหาการกำหนดลักษณะโมเดลคลาดเคลื่อนด้วย

ธรรมชาติของความเป็นพหุมิติที่เหมือนกันของโมเดลในภาพที่ 3.14 กับโมเดลในภาพที่ 3.15 และ 3.16 คือ ความเป็นพหุมิติเป็นคุณสมบัติของข้อมูลที่ได้จากข้อสอบหรือแบบสอบถูกออกแบบมาเพื่อวัดความเป็นพหุมิตินั้น ข้อสอบหรือแบบสอบที่ไม่ได้ออกแบบมาเพื่อวัดสิ่งที่สนใจหลายมิติไปพร้อมกันตั้งแต่ต้นจะไม่สมเหตุสมผลที่จะตรวจสอบว่าข้อมูลที่ได้จากข้อสอบหรือแบบสอบนั้นมีความเป็นพหุมิติหรือไม่ นอกจากนี้ความเป็นพหุมิติที่แตกต่างกันของโมเดลในภาพที่ 3.14 กับโมเดลในภาพที่ 3.15 และ 3.16 คือ โมเดลในภาพที่ 3.14 แสดงความเป็นพหุมิติระดับแบบสอบ (Wang, Wilson, & Adams, 1997) เพราะในทางทฤษฎีข้อสอบแต่ละข้อถูกสร้างเพื่อวัดเพียงสาระการเรียนรู้เดียว และรวมข้อสอบทั้งหมดให้อยู่ในแบบสอบฉบับเดียวกันโดยสาระการเรียนรู้หรือมิตีย่อยเหล่านั้นสัมพันธ์กัน ความเป็นพหุมิติจึงไม่ได้เกิดขึ้นภายในคำตอบที่ได้จากข้อสอบแต่ละข้อ แต่จะเกิดขึ้นจากคำตอบข้อสอบทั้งหมด ในทางตรงกันข้ามโมเดลในภาพที่ 3.15 และ 3.16 แสดงความเป็นพหุมิติระดับข้อสอบ (Wang, Wilson, & Adams, 1997) เพราะคำตอบที่ได้จากข้อสอบแต่ละข้อสามารถบ่งชี้ได้ว่าสัมพันธ์กับสาระการเรียนรู้มากกว่าหนึ่งสาระ เพราะผู้สร้างข้อสอบตั้งใจให้ข้อสอบเหล่านั้นวัดมากกว่าหนึ่งสาระการเรียนรู้

การประเมินความสอดคล้องกลมกลืนของข้อสอบและโมเดล

การตรวจสอบคุณภาพข้อสอบจากความสอดคล้องกลมกลืนของคำตอบข้อสอบแต่ละข้อกับข้อสอบข้ออื่นที่อยู่ที่วัดความสามารถทางวิทยาศาสตร์พิจารณาจากสถิติทดสอบความสอดคล้องแบบถ่วงน้ำหนัก (weighted fit) หรือการทดสอบแบบ infit ด้วยค่า Mean square หรือ MNSQ และค่า t แทนที่จะพิจารณาจากสถิติทดสอบแบบไม่ถ่วงน้ำหนัก (unweighted fit) หรือการทดสอบแบบ outfit เนื่องจากการทดสอบแบบ infit จะแกร่งกว่า outfit ในกรณีที่ข้อมูลคะแนนคำตอบข้อสอบทั้งฉบับมีค่าสุดโต่ง เช่นนักเรียนได้คะแนนน้อยมากๆ และ/หรือได้คะแนนเต็ม และมีจำนวนนักเรียนทำข้อสอบน้อย (Wright, 1984) แต่อย่างไรก็ตามในสถานการณ์ส่วนใหญ่ การทดสอบแบบ infit และ outfit จะให้ค่า MNSQ และค่า t คล้ายคลึงกัน (Adams & Khoo, 1996)

ค่า t บ่งชี้ระดับความผันแปรของข้อมูลคำตอบข้อสอบแต่ละข้อตามทฤษฎีการวัดแบบบราสซ์หรือบ่งชี้ความเป็นเอกพันธ์ (item homogeneity) ของคำตอบข้อสอบทั้งหมดที่กำหนดไว้ในโมเดลการวัดเดียวกัน ค่า t ควรจะอยู่ระหว่าง -2 และ $+2$ (Smith, Schumacker, & Busch, 1995; Wright, Mead, & Bell, 1980) ส่วนค่า MNSQ จะอธิบายความสัมพันธ์ระหว่างค่า t และการจำแนกของข้อสอบ (item discrimination) ในสามกรณี (Wilson, Allen, & Li, 2006) คือ ถ้าค่า MNSQ เท่ากับ 1 หมายความว่าความแปรปรวนที่สังเกตได้ (observed variance) จะเท่ากับความแปรปรวนที่คาดหวัง (expected variance) ของการเบี่ยงเบนของการจำแนกข้อสอบ ซึ่งสามารถตีความได้ว่าโค้งคุณลักษณะข้อสอบ (item characteristic curve) เชิงประจักษ์ที่สร้างขึ้นจากข้อมูลคำตอบข้อสอบมีลักษณะเดียวกับโค้งคุณลักษณะข้อสอบที่คาดหวังที่สร้างจากค่าประมาณความสามารถนักเรียน ค่า MNSQ ที่มากกว่า 1 บ่งชี้ว่าโค้งคุณลักษณะข้อสอบเชิงประจักษ์ชันน้อยกว่าโค้งคุณลักษณะข้อสอบที่คาดหวัง และค่า MNSQ ที่น้อยกว่า 1 บ่งชี้ว่าโค้งคุณลักษณะข้อสอบเชิงประจักษ์ชันมากกว่าโค้งคุณลักษณะข้อสอบที่คาดหวัง เนื่องจากไม่มีเกณฑ์ตายตัวสำหรับประเมินขนาดอิทธิพล (effect size) ของการประมาณค่าพารามิเตอร์ข้อสอบและระบุขนาดความสอดคล้องกลมกลืนคลาดเคลื่อนของข้อสอบ (item misfit) ช่วงของค่า MNSQ ที่ได้จากการทดสอบแบบ infit ระหว่าง .75 ถึง 1.33 ถือว่าเป็นช่วงที่ยอมรับกับโดยทั่วไปว่าโค้งคุณลักษณะข้อสอบเชิงประจักษ์มีลักษณะใกล้เคียงกับโค้งคุณลักษณะข้อสอบที่คาดหวัง (Adams & Khoo, 1996; Wilson, Allen, & Li, 2006)

การเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการวัดต่างๆ ดังที่แสดงในตารางที่ 3.6 สามารถพิจารณาได้จากดัชนีสามตัวได้แก่ Deviance index (G^2), Akaike Information Criterion (AIC) และ Bayesian Information Criterion (BIC) ค่า G^2 ซึ่งคำนวณจาก $-2 \times \log\text{-likelihood}$ (Wilson & Boeck, 2004) ใช้หลักการของการแจกแจงแบบ chi-square เพื่อตรวจสอบคุณสมบัติของโมเดล ค่า G^2 น้อยแสดงว่าโมเดลกลมกลืนกับข้อมูลคะแนนคำตอบข้อสอบ ONET มากกว่าโมเดลที่มีค่า G^2 มาก การเปรียบเทียบความกลมกลืนระหว่างสองโมเดลใช้สถิติทดสอบที่เรียกว่า Likelihood ratio โดยสถิติทดสอบนี้มีค่าเท่ากับผลต่างของค่า G^2 ระหว่างสองโมเดล การเปรียบเทียบผลต่างของค่า G^2 แสดงให้เห็นว่าโค้งการแจกแจงหลักหรือโค้งการแจกแจงกลางที่ไม่เอนเอียงไปในทิศทางใดทิศทางหนึ่ง (asymptotic null distribution) ของค่า likelihood ratio ก็คือการรวมกัน (mixture) ของการแจกแจงแบบ chi-square ที่มาจากสองโมเดลที่ถูกถ่วงน้ำหนักเท่าๆ กันกับ degree of freedom โดย degree of freedom นี้มีค่าเท่ากับผลต่างของจำนวนพารามิเตอร์ที่จะถูกประมาณจากแต่ละโมเดล เมื่อเปรียบเทียบระหว่างโมเดลเอกมิตกับโมเดลพหุมิติแล้วพบว่าค่า Likelihood ratio มีนัยสำคัญทางสถิติ นั่นคือ โมเดลพหุมิตกลมกลืนกับข้อมูลมากกว่าโมเดลแบบเอกมิตอย่างมีนัยสำคัญทางสถิติ

ดัชนีประเมินความกลมกลืนตัวที่สองคือ AIC (Akaike, 1974) ดัชนีนี้บอกปริมาณสารสนเทศที่จะหายไปเมื่อนำโมเดลไปใช้อธิบายปรากฏการณ์ที่เกิดขึ้นจริง นั่นคืออธิบายการแลกเปลี่ยนกัน

ระหว่างค่าความคลาดเคลื่อน (bias) กับค่าความแปรปรวน (variance) ที่อธิบายได้ด้วยโมเดล (Shih, Chang, & Cheng, 2013) ตัวอย่างเช่น มีชุดข้อมูลอยู่ชุดหนึ่งซึ่งสามารถสร้างโมเดลแข่งขันได้มากมาย โมเดลแข่งขันเหล่านี้จะถูกเปรียบเทียบโดยคำนวณค่า likelihood ของแต่ละโมเดล เมื่อค่า likelihood มีค่ามาก ก็เท่ากับว่าจะมีค่า G^2 น้อย นั่นคือโมเดลนั้นจะกลมกลืนกับข้อมูลมากกว่า (Janssen & Boeck, 1999) และเมื่อนำโมเดลแข่งขันเหล่านี้มาจัดอันดับตามค่าดัชนี AIC ก็พบว่า โมเดลใดที่มีค่า AIC ต่ำที่สุดแสดงว่าเป็นโมเดลที่กลมกลืนกับข้อมูลมากที่สุด ค่า AIC คำนวณได้จากสมการ (1) เมื่อ p คือจำนวนพารามิเตอร์ที่ถูกประมาณจากโมเดล

$$AIC = -2 \times (\log - \text{likelihood}) + 2p \quad (1)$$

ดัชนีวัดความกลมกลืนตัวที่สาม คือ BIC (Schwarz, 1978) โดยดัชนีนี้ให้น้ำหนักของกลุ่มตัวอย่างมาคำนวณร่วมด้วย แทนที่จะมีเฉพาะจำนวนพารามิเตอร์ที่ต้องประมาณเพียงอย่างเดียว ดัชนี BIC สามารถคำนวณได้จากสมการ (2) เมื่อ n คือขนาดกลุ่มตัวอย่าง

$$BIC = -2 \times (\log - \text{likelihood}) + p(\log n) \quad (2)$$

ดัชนี BIC เป็นการปรับแก้ดัชนี AIC ในสถานการณ์ที่มีพารามิเตอร์จำนวนมากที่ต้องประมาณ ซึ่งส่งผลให้โมเดลนั้นเป็นโมเดลที่ดีกว่าเมื่อเทียบกับโมเดลอื่นๆ ที่มีจำนวนพารามิเตอร์น้อยกว่า ข้อสรุปเกี่ยวกับการเลือกใช้ดัชนี AIC และ BIC เพื่อประเมินความกลมกลืนของโมเดลคือ โมเดลที่มีค่า AIC และ BIC น้อยที่สุดเป็นโมเดลที่กลมกลืนกับผลการตอบข้อสอบมากที่สุด แต่อย่างไรก็ตามดัชนีทั้งสองตัวนี้ไม่จำเป็นที่จะต้องเป็นไปในทิศทางเดียวกันเสมอไป (Lin & Dayton, 1997)

ขั้นตอนการวิเคราะห์คะแนนจุดตัด เริ่มต้นหลังจากได้โมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด การใช้ค่าประมาณความสามารถนักเรียน (θ) โดยตรงเพื่อตีความเป็นคุณลักษณะนักเรียน หรือสรุปอ้างอิงเกี่ยวกับความสามารถนักเรียนในสถานการณ์ทั่วไปอาจจะไม่สะดวกและอาจจะทำให้ผู้ใช้ประโยชน์จากค่าประมาณดังกล่าวเข้าใจคลาดเคลื่อนได้ ดังนั้นจึงแปลงค่าประมาณความสามารถนักเรียนและค่าความยากข้อสอบซึ่งอยู่บนสเกลเดียวกัน โดยมีค่าเฉลี่ยเท่ากับ 50 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 ตามสมการ (3)

$$\theta^* = 10(\theta) + 50 \quad (3)$$

ค่าแปลงความสามารถนักเรียนสามารถนำไปใช้ได้หลากหลายในขั้นตอนกำหนดคะแนนจุดตัด ตัวอย่างเช่น จากภาพที่ 3.17 ซึ่งเป็นตัวอย่างสมมติของแผนที่ของ Wright ที่ได้จากการวิเคราะห์โมเดลเอกมิติของแบบสอบ ONET สมมติว่านักวิจัยสนใจที่จะกำหนดคะแนนจุดตัดที่ 40 คะแนน ($\theta = -1.0$) ก็จะเห็นว่านักเรียนที่มีคะแนนตั้งแต่ 40 คะแนนขึ้นไปมีโอกาสน้อย 50%

ที่จะตอบถูกข้อสอบ ONET ส่วนใหญ่ ค่าความน่าจะเป็นที่อยู่ในคอลัมน์ซ้ายสุดเป็นความน่าจะเป็นของนักเรียนที่มีคะแนนเท่ากับ 40 ที่จะตอบข้อสอบแต่ละข้อได้ถูกต้อง นั่นคือนักเรียนมีโอกาส 60% ที่จะตอบข้อที่ 34, 43, 45, และ 48 ได้ถูก มีโอกาส 39% ที่จะตอบข้อที่ 37 ได้ถูก และมีโอกาส 86% ที่จะตอบข้อที่ 2 ได้ถูก นอกจากนี้ นักวิจัยอาจจะกำหนดคะแนนจุดตัดจากข้อสอบจำนวนครึ่งหนึ่งที่นักเรียนมีโอกาสตอบถูกอย่างน้อย 50% ซึ่งตรงกับ 36 คะแนนก็ได้ เป็นต้น

คะแนน	ความยากข้อสอบ	ความน่าจะเป็นที่จะตอบถูกของนักเรียนที่มีคะแนนเท่ากับ 40 (P@40)
55		
54		
53		
52		
51		
50		
49		
48	37	.39
47	15	.44
46		
45	28 39	.48
44	27	.51
43	38	.53
42	19	.55
41		
@40	34 43 45 48	.60
39	17 18 20 40 50	.62
38	31	.65
37	4 11 33 44	.66
36	5 9 12 32 46 47	.69
35	3 7 16	.71
34	6 10 29	.72
33	22 35 36	.75
32	8 14 23 26	.76
31	13 24 25	.78
30	41 42	.79
29	1 21 49	.81
28	30	.82
27		
26		
25	2	.86

ภาพที่ 3.17 ตัวอย่างการกำหนดคะแนนจุดตัดด้วยวิธีสร้างแผนที่ตัวแปรเชิงทฤษฎี

แนวทางในการกำหนดคะแนนจุดตัดข้างต้นทำได้ทั้งแบบอิงเกณฑ์และอิงกลุ่ม ตัวอย่างที่นำเสนอข้างต้นเป็นแนวทางแบบอิงเกณฑ์ สำหรับแนวทางแบบอิงกลุ่มในเบื้องต้น อาจจะพิจารณาจากค่าเฉลี่ยคะแนนสอบที่ได้ตามทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) เมื่อนำหลักการกำหนดคะแนนจุดตัดด้วยวิธีทำแผนที่ภาวะสันนิษฐานเชิงประจักษ์กับโมเดลพหุมิติที่กลมกลืนกับผลการตอบข้อสอบของนักเรียนก็จะพบว่า สามารถกำหนดคะแนนจุดตัดได้อย่างมีประสิทธิภาพมากขึ้น และให้สารสนเทศเชิงวินิจัยที่ละเอียดมากขึ้นอีกด้วย เมื่อพิจารณาตัวอย่างสมมติของแผนที่ของ Wright ของแต่ละปีการศึกษาที่นำมาเรียงต่อกันดังภาพที่ 3.18 จะพบแนวโน้ม

ของความสามารถของนักเรียนแต่ละปี รวมทั้งช่วยชี้แนะข้อมูลสำหรับกำหนดคะแนนจุดตัดร่วม (common cutscore) ที่เหมาะสมได้

คะแนน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	ความยากข้อสอบ	P@40	ความยากข้อสอบ	P@40	ความยากข้อสอบ	P@40
60						
59						
58						
57			8	.10		
56			28 34	.12		
55						
54			35 48	.15		
53			9 36	.18		
52			27	.19	37	.39
51			45	.20	15	.44
50			4 26	.21	39	.45
49			44	.23	28	.48
48	37	.39	3 29	.25	27	.51
47	15	.44	40 49	.26	38	.53
46			30 37	.28	19	.55
45	28 39	.48	14 43 50	.29		
44	27	.51	5 31 41	.32	34 43 45 48	.60
43	38	.53	1 6 10 15	.35	17 18 20 40 50 31	.62
42	19	.55	16 17 18	.37	2	.65
41			32 33 42	.39	4 11 33 44	.66
@40	34 43 45 48	.60	11 19 20	.40	5 9 12 32 46 47	.69
39	17 18 20 40 50	.62	21 38	.43	3 7 16	.71
38	31	.65	39 47	.45	6 10 29	.72
37	4 11 33 44	.66	7 46	.48	22 35 36	.75
36	5 9 12 32 46 47	.69	22	.50	8 14 23 26	.76
35	3 7 16	.71	2 25	.52	13 24 25	.78
34	6 10 29	.72			41 42	.79
33	22 35 36	.75	13	.56	1 21 49	.81
32	8 14 23 26	.76			30	.82
31	13 24 25	.78	23	.60		
30	41 42	.79	12 24	.63		
29	1 21 49	.81				
28	30	.82				
27						
26						
25	2	.86				

ภาพที่ 3.18 ตัวอย่างการกำหนดคะแนนจุดตัดระหว่างปีการศึกษา

ตอนที่ 2 ผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์

โครงการวิจัยนี้ใช้โมเดลเอกมิติ (โมเดล UniCM และโมเดล ConsUniCM ในตารางที่ 3.6) เพื่อตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์ของนักเรียน รายละเอียดการเรียนรู้ (ภาพที่ 3.3 ถึง 3.10) และแบบภาพรวม (ภาพที่ 3.11) ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลคำตอบข้อสอบ ONET จากกลุ่มตัวอย่างขนาด 5% ของฐานข้อมูลวิเคราะห์จะช่วยประเมินว่าโครงสร้างความสามารถทางวิทยาศาสตร์ของนักเรียนตามที่

ผู้เชี่ยวชาญได้เสนอแนะไว้จำนวนห้าระดับเหมาะสมหรือไม่ และนำไปสู่การแก้ไขปรับปรุงแผนที่ภาวะสันนิษฐานเชิงทฤษฎีให้เป็นแผนที่ภาวะสันนิษฐานเชิงประจักษ์สำหรับตีความคะแนนจุดตัดขั้นต้น

หลักเกณฑ์การพิจารณาความเชื่อมโยงกันระหว่างระดับความสามารถตามผู้เชี่ยวชาญกำหนดขึ้นและระดับความยากของข้อสอบที่ได้จากแผนที่ของ Wright นั้น ไม่สามารถพิจารณาได้จากค่าพารามิเตอร์ความยากข้อสอบได้โดยตรง เนื่องจากข้อสอบ ONET ของแต่ละปีการศึกษาเป็นคนละฉบับและไม่มีหลักฐานแสดงความสัมพันธ์กันของแบบสอบแต่ละปีการศึกษาด้วย นอกจากนี้ นักเรียนที่ทำข้อสอบ ONET แต่ละปีการศึกษาก็มาจากประชากรคนละกลุ่ม อีกทั้งเกณฑ์พิจารณาระดับความยากข้อสอบยังขึ้นอยู่กับกรอบอ้างอิง (frame of reference) ที่ใช้ (Baker, 2001) ดังนั้นเมื่อกรอบอ้างอิงเปลี่ยนไปข้อสอบยากอาจกลายเป็นข้อสอบง่ายของกรอบอ้างอิงใหม่ ซึ่งโครงการวิจัยนี้ใช้กรอบอ้างอิงความยากข้อสอบที่สัมพันธ์กับระดับความสามารถนักเรียนจากผลประเมินโดยผู้เชี่ยวชาญในภาพที่ 3.2 หลักการพิจารณาความยากข้อสอบที่ได้จากแผนที่ของ Wright จึงไม่ได้อยู่ที่ค่าพารามิเตอร์ความยากข้อสอบเป็นหลักแต่ขึ้นอยู่กับอันดับความยากข้อสอบและความหมายของความยากข้อสอบ (ที่สัมพันธ์กับความสามารถนักเรียน) ที่ใช้บรรยายความผันแปรความยากข้อสอบและความสามารถนักเรียนที่เกิดขึ้นในแต่ละปีการศึกษา รายละเอียดของผลการตรวจสอบคุณภาพของแผนที่ภาวะสันนิษฐานเชิงทฤษฎีมีดังนี้

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

เมื่อพิจารณาแผนที่ของ Wright ที่ได้จากวิธีวิเคราะห์เอกมิติแบบแยกรายมิติโดยใช้ข้อมูลคำตอบข้อสอบ ONET ที่วัดเฉพาะสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิตของปีการศึกษา 2550 ถึง 2552 (ดูภาพที่ A5.1 ในภาคผนวก) พบว่าในภาพรวมตำแหน่งค่าความยากข้อสอบแต่ละปีจะอยู่สูงกว่าค่าความสามารถนักเรียน แสดงว่าข้อสอบส่วนใหญ่เป็นข้อสอบค่อนข้างยากเมื่อเปรียบเทียบกับความสามารถนักเรียนของปีนั้น และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV (EAP/PV reliability) ซึ่งเท่ากับ .44, .45, .47 และ .45 (เรียงลำดับจากภาพที่ A5.1a ถึง A5.1d) พบว่ามีค่าปานกลางในระดับเดียวกันแสดงว่าความแปรปรวนของความสามารถนักเรียนที่ทำข้อสอบในแต่ละปีถูกอธิบายด้วยข้อสอบได้เพียงบางส่วน

เมื่อพิจารณาดำแหน่งความยากข้อสอบในแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.3) พบว่าลำดับความยากข้อสอบส่วนใหญ่สอดคล้องเป็นไปในทิศทางเดียวกัน ยกเว้นข้อ 50.1, 50.7, 51.4, 51.5, 52.3, 52.4 และ 52.6 ที่มีลำดับความยากข้อสอบผันแปรสลับข้ามขั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นจึงเป็นไปได้ว่าความสามารถนักเรียนในชั้นเชี่ยวชาญและขั้นพื้นฐานอาจจะปะปนกัน และเมื่อพิจารณาให้ทั้งสองชั้นเป็นชั้นเดียวกันพบว่า แผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถ

บรรยายลำดับความยากข้อสอบสอดคล้องกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีและแผนที่ของ Wright ได้ดีขึ้น

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

เมื่อพิจารณาแผนที่ของ Wright จากโมเดลเอกมิตีที่ประกอบด้วยข้อสอบที่วัดเฉพาะสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อมของปีการศึกษา 2550 ถึง 2552 (ดูภาพที่ A5.2 ในภาคผนวก) พบว่าในภาพรวมตำแหน่งค่าความยากข้อสอบทั้งสามปีจะอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียน แสดงให้เห็นว่าข้อสอบทั้งสามปีเป็นข้อสอบยากปานกลางสำหรับนักเรียนในปีการศึกษานั้น เมื่อพิจารณาดำเนินค่าความยากข้อสอบปีการศึกษา 2552 พบว่ามีตำแหน่งค่อนข้างสูงกว่าค่าความสามารถนักเรียนส่วนใหญ่ และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .27, .30, .23 และ .25 (เรียงลำดับจากภาพที่ A5.2a ถึง A5.2d) พบว่ามีค่าต่ำซึ่งแสดงว่าความแปรปรวนของความสามารถนักเรียนยังไม่ถูกอธิบายได้ดีเพียงพอด้วยข้อสอบของแต่ละปี

เมื่อพิจารณาลำดับความยากข้อสอบในแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.4) พบว่ามีลำดับความยากข้อสอบส่วนใหญ่สอดคล้องกัน แต่อย่างไรก็ตามข้อสอบข้อ 50.8, 51.8 และ 51.9 มีลำดับความยากข้อสอบสลับข้ามขึ้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎีทำให้แผนที่ภาวะสันนิษฐานเชิงทฤษฎีไม่สามารถอธิบายความแปรปรวนในลำดับความยากข้อสอบในแผนที่ของ Wright นอกจากนี้มีข้อสอบวัดสาระการเรียนรู้จำนวนน้อยซึ่งเป็นสาเหตุที่ทำให้ลำดับค่าความยากข้อสอบในแผนที่ของ Wright ไม่คงเส้นคงวา (Wright, 1992) ดังนั้นผู้วิจัยจึงเชื่อว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์มีลักษณะเดียวกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่สามารถใช้บรรยายลำดับลำดับความยากข้อสอบส่วนใหญ่ได้

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้สารและสมบัติของสาร

เมื่อพิจารณาจากแผนที่ของ Wright (ดูภาพที่ A5.3 ในภาคผนวก) ที่ได้จากวิธีวิเคราะห์เอกมิตีเฉพาะข้อสอบที่วัดสาระการเรียนรู้สารและสมบัติของสารพบว่า ในปีการศึกษา 2550 ตำแหน่งค่าความยากข้อสอบอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียนแสดงว่าข้อสอบปีการศึกษา 2550 เป็นข้อสอบยากปานกลาง สำหรับปีการศึกษา 2551 พบว่าตำแหน่งค่าความยากข้อสอบส่วนใหญ่จะอยู่ต่ำกว่าการแจกแจงค่าความสามารถนักเรียนแสดงว่าข้อสอบเป็นข้อสอบง่าย และปีการศึกษา 2552 พบว่าตำแหน่งค่าความยากข้อสอบอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่ แสดงให้เห็นว่าเป็นข้อสอบค่อนข้างยาก และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .58, .58 และ .46 (เรียงลำดับจากภาพที่ A5.3a ถึง A5.3c) พบว่ามีปานกลางในระดับเดียวกัน แสดงว่าความแปรปรวนของความสามารถนักเรียนถูกอธิบายด้วยข้อสอบในแต่ละปีการศึกษาเพียงบางส่วน

เมื่อพิจารณาลำดับความยากข้อสอบจากแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.5) พบว่าลำดับความยากข้อสอบส่วนใหญ่สอดคล้องกัน แต่อย่างไรก็ตามมีข้อสอบข้อ 50.14, 50.17, 51.14 และ 51.16 ที่มีลำดับความยากข้อสอบสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นจึงเป็นไปได้ว่าความสามารถนักเรียนในชั้นพื้นฐานและขั้นต้นอาจจะปะปนกัน และเมื่อพิจารณาให้ทั้งสองระดับเป็นชั้นความสามารถเดียวกันจะพบว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถบรรยายลำดับความยากข้อสอบสอดคล้องกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีและแผนที่ของ Wright ได้ดีขึ้น

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้แรงและการเคลื่อนที่

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.4 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติเฉพาะข้อสอบที่วัดสาระการเรียนรู้แรงและการเคลื่อนที่ พบว่าตำแหน่งค่าความยากข้อสอบปีการศึกษา 2550 และ 2551 ส่วนใหญ่จะอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียนแสดงว่าเป็นข้อสอบยากปานกลาง สำหรับปีการศึกษา 2552 พบว่าตำแหน่งค่าความยากข้อสอบอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่แสดงเป็นข้อสอบค่อนข้างยาก และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .38, .33, .36 และ .32 (เรียงลำดับจากภาพที่ A5.4a ถึง A5.4d) พบว่ามีค่าต่ำในระดับเดียวกันซึ่งแสดงว่าความแปรปรวนของความสามารถนักเรียนของแต่ละปีการศึกษาถูกอธิบายได้อย่างเพียงบางส่วนเท่านั้น

เมื่อพิจารณาลำดับค่าความยากข้อสอบส่วนใหญ่จากแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.6) พบว่ามีลำดับความยากข้อสอบสอดคล้องไปในทิศทางเดียวกัน ยกเว้นข้อ 51.21 และ 52.7 ที่ข้อสอบมีลำดับความยากสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นจึงเป็นไปได้ที่ความสามารถนักเรียนในชั้นพื้นฐานและขั้นต้นอาจจะปะปนกัน และเมื่อพิจารณาให้ทั้งสองระดับเป็นชั้นเดียวกันพบว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถบรรยายลำดับความยากข้อสอบสอดคล้องกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีและแผนที่ของ Wright ได้ดีขึ้น

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้พลังงาน

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.5 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติเฉพาะข้อสอบที่วัดสาระการเรียนรู้พลังงาน ในปีการศึกษา 2550 และ 2551 พบว่าตำแหน่งค่าความยากข้อสอบส่วนใหญ่อยู่บริเวณตรงกลางการแจกแจงค่าความสามารถนักเรียนของแต่ละปี แสดงว่าเป็นข้อสอบยากปานกลาง สำหรับปีการศึกษา 2552 ตำแหน่งค่าความยากข้อสอบจะอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่แสดงว่าเป็นข้อสอบค่อนข้างยาก และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .45, .43, .34 และ .31 (เรียงลำดับจากภาพที่ A5.5a ถึง A5.5d) พบว่ามีค่าต่ำใน

ระดับเดียวกัน แสดงว่าข้อสอบ ONET ในแต่ละปีการศึกษานิยามความแปรปรวนของความสามารถนักเรียนได้เพียงบางส่วน

เมื่อพิจารณาลำดับค่าความยากข้อสอบในแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.7) พบว่าข้อสอบส่วนใหญ่มีลำดับความยากข้อสอบสอดคล้องไปในทิศทางเดียวกัน ยกเว้นข้อ 50.29 และ 51.23 ที่ข้อสอบมีลำดับความยากสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นผู้วิจัยจึงเชื่อว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์มีลักษณะเดียวกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่สามารถใช้บรรยายลำดับค่าความยากข้อสอบส่วนใหญ่ได้

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.6 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติเฉพาะข้อสอบที่วัดสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก พบว่าในภาพรวมตำแหน่งค่าความยากข้อสอบปีการศึกษา 2550 จะอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียน แสดงว่าเป็นข้อสอบยากปานกลาง เมื่อพิจารณาปีการศึกษา 2551 และ 2552 พบว่าตำแหน่งค่าความยากข้อสอบอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่แสดงว่าเป็นข้อสอบค่อนข้างยาก และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .43, .42, และ .27 (เรียงลำดับจากภาพที่ A5.6a ถึง A5.6c) พบว่ามีค่าต่ำ ซึ่งแสดงว่าข้อสอบ ONET ที่วัดสาระการเรียนรู้นี้อธิบายความแปรปรวนของความสามารถนักเรียนได้เพียงบางส่วน

เมื่อพิจารณาลำดับค่าความยากข้อสอบจากแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.8) พบว่าข้อสอบมีลำดับความยากสอดคล้องในทิศทางเดียวกัน ยกเว้นข้อ 50.32 และ 51.33 ซึ่งมีลำดับความยากข้อสอบสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นผู้วิจัยจึงเชื่อว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์มีลักษณะเดียวกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่สามารถใช้บรรยายลำดับค่าความยากข้อสอบส่วนใหญ่ได้

2.7 แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ดาราศาสตร์และอวกาศ

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.7 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติเฉพาะข้อสอบที่วัดสาระการเรียนรู้ดาราศาสตร์และอวกาศ พบว่า ในภาพรวมตำแหน่งค่าความยากข้อสอบปีการศึกษา 2550 และ 2551 จะอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียน แสดงว่าข้อสอบปีการศึกษาดังกล่าวเป็นข้อสอบยากปานกลาง สำหรับปีการศึกษา 2552 พบว่าตำแหน่งค่าความยากข้อสอบอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่ แสดงว่าเป็นข้อสอบค่อนข้างยาก และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .45, .48, .26 และ .26

(เรียงลำดับจากภาพที่ A5.7a ถึง A5.7d) พบว่ามีค่าต่ำในระดับเดียวกัน นั่นหมายความว่าข้อสอบ ONET ของแต่ละปีการศึกษาอธิบายความแปรปรวนของความสามารถนักเรียนได้น้อยและมีความแปรปรวนที่เหลือน้อยส่วนใหญ่ที่ข้อสอบดังกล่าวไม่สามารถอธิบายได้

เมื่อพิจารณาลำดับความยากข้อสอบจากแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.9) พบว่าลำดับความยากข้อสอบส่วนใหญ่สอดคล้องในทิศทางเดียวกัน ยกเว้นข้อ 51.40 ที่มีลำดับความยากข้อสอบผันแปรสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นผู้วิจัยจึงเชื่อว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์มีลักษณะเดียวกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีที่สามารถบรรยายลำดับความยากข้อสอบส่วนใหญ่ได้

แผนที่ภาวะสันนิษฐานเชิงประจักษ์สาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.8 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติเฉพาะข้อสอบที่ผู้เชี่ยวชาญประเมินว่าวัดสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี พบว่าในภาพรวมตำแหน่งค่าความยากข้อสอบแต่ละปีจะอยู่บริเวณตรงกลางของการแจกแจงค่าความสามารถนักเรียน แสดงว่าข้อสอบส่วนใหญ่ในแต่ละปีการศึกษาเป็นข้อสอบยากปานกลาง และเมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .42, .53, และ .53 (เรียงลำดับจากภาพที่ A5.8a ถึง A5.8c) พบว่ามีค่าปานกลางในระดับเดียวกันซึ่งแสดงว่าความแปรปรวนของความสามารถนักเรียนที่ทำข้อสอบในแต่ละปีการศึกษายังไม่ถูกอธิบายอย่างเพียงพอ

เมื่อพิจารณาลำดับความยากข้อสอบจากแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (ดูภาพที่ 3.10) พบว่าข้อสอบส่วนใหญ่มีลำดับความยากข้อสอบสอดคล้องกัน ยกเว้นข้อ 50.8, 51.14, 51.21, 51.28, 52.3, 52.4 และ 52.7 ที่ลำดับความยากข้อสอบสลับข้ามชั้นระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี ดังนั้นจึงเป็นไปได้ว่าความสามารถนักเรียนในระดับชั้นสูงและชั้นเชี่ยวชาญอาจจะปะปนกัน และเมื่อพิจารณาให้ทั้งสองระดับเป็นชั้นเดียวกัน พบว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถบรรยายลำดับความยากข้อสอบสอดคล้องกับแผนที่ภาวะสันนิษฐานเชิงทฤษฎีและแผนที่ของ Wright ได้ดีขึ้น

แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์แบบภาพรวมและข้อสรุป

เมื่อพิจารณาแผนที่ของ Wright (ดูภาพที่ A5.9 ในภาคผนวก) ที่ได้จากโมเดลเอกมิติของข้อสอบทุกข้อในแต่ละปีการศึกษา พบว่าในภาพรวมตำแหน่งค่าความยากข้อสอบจะอยู่สูงกว่าค่าความสามารถนักเรียนส่วนใหญ่ แสดงให้เห็นว่าข้อสอบในปีการศึกษา 2550 และ 2551 เป็นข้อสอบค่อนข้างยาก และข้อสอบในปีการศึกษา 2552 เป็นข้อสอบยาก เมื่อพิจารณาค่าความเที่ยงแบบ EAP/PV ซึ่งเท่ากับ .83, .83, .77 และ .75 (เรียงลำดับจากภาพที่ A5.9a ถึง A5.9d) พบว่ามีค่าสูงใน

ระดับเดียวกัน แสดงว่าข้อสอบ ONET ในแต่ละปีการศึกษาสามารถอธิบายความแปรปรวนของความสามารถนักเรียนได้ดี (เนื่องจากมีจำนวนข้อสอบมากขึ้นสำหรับใช้บรรยายโครงสร้างความสามารถนักเรียนได้ดีขึ้น)

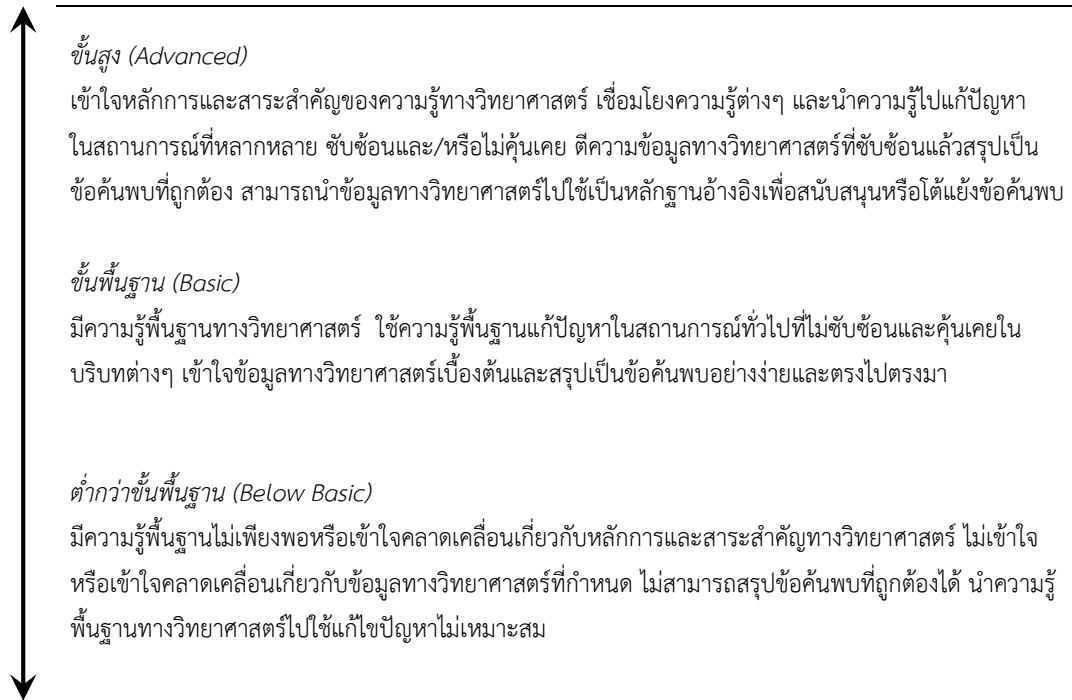
เมื่อพิจารณาลำดับความยากข้อสอบในปีการศึกษา 2550, 2551 และ 2552 จำนวนทั้งหมด 117 ข้อพบว่าข้อสอบส่วนใหญ่มีลำดับความยากสอดคล้องกันระหว่างลำดับความยากในแผนที่ของ Wright ซึ่งเป็นหลักฐานเชิงประจักษ์ และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี (แผนภาพ 3.11) ซึ่งเป็นผลการประเมินโดยผู้เชี่ยวชาญ กล่าวคือมีข้อสอบที่มีลำดับความยากตรงกันกับระดับขั้นเชี่ยวชาญ จำนวน 6 ข้อหรือร้อยละ 5.13 ระดับขั้นพื้นฐานจำนวน 28 ข้อหรือร้อยละ 23.93 และระดับขั้นต้นจำนวน 9 ข้อหรือร้อยละ 7.69 ของข้อสอบทั้งหมดดังตารางที่ 3.7

ตารางที่ 3.7 ความสอดคล้องของลำดับความยากข้อสอบระหว่างแผนที่ของ Wright และระดับความสามารถนักเรียนในแผนที่ภาวะสันนิษฐานเชิงทฤษฎี

ความสอดคล้อง	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552		รวม	
	จำนวนข้อ	ร้อยละ	จำนวนข้อ	ร้อยละ	จำนวนข้อ	ร้อยละ	จำนวนข้อ	ร้อยละ
สูง-เชี่ยวชาญ	2	5.00	-	-	-	-	2	1.71
สูง-พื้นฐาน	-	-	3	7.50	2	5.41	5	4.27
สูง-ต้น	-	-	-	-	-	-	-	-
สูง-ต่ำกว่าขั้นต้น	-	-	-	-	-	-	-	-
เชี่ยวชาญ-สูง	-	-	1	2.50	4	10.81	5	4.27
เชี่ยวชาญ-พื้นฐาน	6	15.00	-	-	1	2.70	7	5.98
เชี่ยวชาญ-ต้น	2	5.00	-	-	2	5.41	4	3.42
เชี่ยวชาญ-ต่ำกว่าขั้นต้น	-	-	-	-	-	-	-	-
พื้นฐาน-สูง	1	2.50	4	10.00	-	-	5	4.27
พื้นฐาน-เชี่ยวชาญ	1	2.50	4	10.00	3	8.11	9	7.69
พื้นฐาน-ต้น	2	5.00	3	7.50	2	5.41	7	5.98
พื้นฐาน-ต่ำกว่าขั้นต้น	-	-	-	-	-	-	-	-
ต้น-สูง	-	-	2	-	-	-	2	1.71
ต้น-เชี่ยวชาญ	4	10.00	3	7.50	1	2.70	8	6.84
ต้น-พื้นฐาน	5	12.50	4	12.50	10	27.03	20	17.09
ต้น-ต่ำกว่าขั้นต้น	-	-	-	-	-	-	-	-
คงที่ขั้นสูง	-	-	-	-	-	-	-	-
คงที่ขั้นเชี่ยวชาญ	1	2.50	5	12.50	1	2.70	6	5.13
คงที่ขั้นพื้นฐาน	12	30.00	8	20.00	8	21.62	28	23.93
คงที่ขั้นต้น	4	10.00	3	7.50	3	8.11	9	7.69
คงที่ขั้นต่ำกว่าขั้นต้น	-	-	-	-	-	-	-	-
รวม	40	100	40	100	37	100	117	100

เมื่อพิจารณาความผันแปรของลำดับความยากข้อสอบระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐานเชิงทฤษฎี พบว่าเกิดขึ้นกับชั้นความสามารถที่อยู่ติดกันและแบบข้ามชั้น ได้แก่ (1) ชั้นสูงสลับกับชั้นเชี่ยวชาญจำนวน 7 ข้อหรือร้อยละ 5.98 (2) ชั้นสูงสลับกับชั้นพื้นฐานจำนวน 10 ข้อหรือร้อยละ 8.55 (3) ชั้นสูงสลับกับชั้นต้นจำนวน 2 ข้อหรือร้อยละ 1.71 (3) ชั้นเชี่ยวชาญสลับกับชั้นพื้นฐานจำนวน 16 ข้อหรือร้อยละ 13.68 (4) ชั้นเชี่ยวชาญสลับกับชั้นต้นจำนวน 12 ข้อหรือร้อยละ 10.26 และ (5) ชั้นพื้นฐานสลับกับชั้นต้นจำนวน 27 ข้อหรือร้อยละ 23.08 นอกจากนี้จะเห็นว่าลำดับความยากข้อสอบส่วนใหญ่คงเส้นคงวา และจะอธิบายความแปรปรวนของลำดับความยากข้อสอบและความสามารถนักเรียนให้ดียิ่งขึ้นเมื่อยุบรวมชั้นพื้นฐานกับชั้นต้นเข้าด้วยกัน และเมื่อนำผลวิเคราะห์ที่ได้จากแผนที่ภาวะสันนิษฐานเชิงประจักษ์รายละเอียดการเรียนรู้มาพิจารณาร่วมกัน จะพบว่าเมื่อยุบรวมกันแล้วข้อสอบในชั้นที่เกิดขึ้นใหม่จะมีจำนวนเพิ่มขึ้นเป็น 64 ข้อ (27 ข้อ+28 ข้อ+9 ข้อ) คิดเป็นร้อยละ 54.70 จากข้อสอบทั้งหมด 117 ข้อ นอกจากนี้เมื่อพิจารณาข้อสอบปีการศึกษา 2552 พบว่าเป็นข้อสอบยาก ซึ่งอาจเกิดขึ้นจากรูปแบบข้อสอบที่กำหนดให้นักเรียนเลือกคำตอบสองตัวเลือก จึงทำให้ลำดับความยากข้อสอบข้อ 33 ถึง 37 ยากกว่าชั้นเชี่ยวชาญ ดังนั้นผู้วิจัยจึงยุบรวมชั้นสูงและชั้นเชี่ยวชาญเข้าด้วยกันเพื่อให้แผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถบรรยายความสามารถนักเรียนได้ทุกสาระการเรียนรู้และทุกปีการศึกษา (ถ้าไม่ยุบรวมจะใช้บรรยายความสามารถนักเรียนปีการศึกษา 2552 ไม่ได้) และสำหรับความผันแปรลำดับความยากข้อสอบแบบข้ามชั้น ผู้วิจัยจะกำหนดให้ข้อสอบข้อนั้นมีลำดับความยากตามแผนที่ของ Wright ซึ่งเป็นข้อมูลเชิงประจักษ์แทนที่จะกำหนดตามความคิดเห็นของผู้เชี่ยวชาญ และถือว่าความผันแปรที่เกิดขึ้นนี้เป็นความแปรปรวนที่ไม่สามารถอธิบายได้ด้วยแผนที่ภาวะสันนิษฐานเชิงประจักษ์

ผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานของแต่ละสาระการเรียนรู้ที่ผู้วิจัยเสนอแนะให้ยุบรวมชั้นสูงและชั้นเชี่ยวชาญและยุบรวมชั้นพื้นฐานกับชั้นต้น หรือแม้กระทั่งไม่ยุบรวมระดับชั้นความสามารถใดเลยทำให้เกิดรูปแบบการตีความคะแนนสอบ ONET หลากหลายและอาจจะสับสนมากขึ้นเมื่อนำไปใช้กำหนดคะแนนจุดตัดขั้นต้น ผู้วิจัยจึงสรุปให้แผนที่ภาวะสันนิษฐานเชิงประจักษ์ของแต่ละสาระการเรียนรู้และภาพรวมมีรูปแบบการตีความความสามารถของนักเรียนแบบเดียวกัน โดยจะยุบรวมชั้นสูงและชั้นเชี่ยวชาญเข้าด้วยกันและเรียกชั้นนี้ว่า ชั้นสูง และยุบรวมชั้นพื้นฐานและชั้นต้นเข้าด้วยกันและเรียกชั้นนี้ว่า ชั้นพื้นฐาน โดยหลักการที่ผู้วิจัยใช้ยุบรวมชั้นความสามารถนักเรียนดังกล่าวจะต้องไม่ขัดแย้งกับผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานรายละเอียดการเรียนรู้และในภาพรวม ซึ่งแม้ว่าอาจจะทำให้ผลการวิจัยสูญเสียสารสนเทศบางส่วนไป แต่จะช่วยให้แผนที่ภาวะสันนิษฐานเชิงประจักษ์สามารถบรรยายความสามารถนักเรียนได้คงเส้นคงวามากขึ้นและตีความคะแนนสอบ ONET ที่สัมพันธ์กับคะแนนจุดตัดขั้นต้นสะดวกมากขึ้น ดังนั้นแผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 ทั้งแบบรายละเอียดการเรียนรู้ที่ 1 ถึง 8 และในภาพรวมจึงประกอบด้วยระดับความสามารถ 3 ชั้นได้แก่ชั้นต่ำกว่าชั้นพื้นฐาน ชั้นพื้นฐานและชั้นสูง ดังแสดงในภาพที่ 3.19



ภาพที่ 3.19 แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์

บทที่ 4

ผลการวิจัย

วัตถุประสงค์หลักของโครงการวิจัยนี้คือเพื่อวิเคราะห์ ตรวจสอบและนำเสนอคะแนนจุดตัดขั้นต้น (tentative cutscore) ความสามารถทางวิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากคะแนนสอบระดับชาติขั้นพื้นฐาน (Ordinary National Educational Test: ONET) ระหว่างปีการศึกษา 2550 ถึง 2552 ผลการวิจัยบางส่วนเกี่ยวกับการพัฒนาแผนที่ภาวะสันนิษฐานเชิงประจักษ์สำหรับใช้แสดงโครงสร้างความสามารถนักเรียนและบรรยายคุณลักษณะนักเรียนได้นำเสนอแล้วในตอนที่ 2 ของบทที่ 3 เนื้อหาในบทนี้จึงนำเสนอผลการวิจัยส่วนที่เหลือซึ่งสัมพันธ์กับส่วนประกอบที่ 4 ถึง 6 ของกรอบแนวคิดการวิจัยและครอบคลุมผลวิเคราะห์ต่างๆ จากกระบวนการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลคำตอบข้อสอบ ONET การตรวจสอบความตรงข้ามกลุ่มของโมเดล การสร้างสเกลคะแนนความสามารถทางวิทยาศาสตร์ การกำหนดคะแนนจุดตัดขั้นต้นและการวิเคราะห์ผลกระทบที่เกิดขึ้น รายละเอียดของผลการวิจัยมีรายละเอียดดังต่อไปนี้

ตอนที่ 1 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลการวัด

ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2550

ตามที่ได้อธิบายไว้แล้วในบทที่ 3 โมเดลการวัดที่ถูกพัฒนาขึ้นมีหลักการพื้นฐานตามโครงสร้างความสามารถทางวิทยาศาสตร์ที่กำหนดไว้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 และตามข้อกำหนดลักษณะแบบสอบ ONET ประกอบด้วยโมเดลเอกมิติ (unidimensional model) และโมเดลพหุมิติระหว่างข้อสอบ (between-item multidimensional model) และโมเดลพหุมิติภายในข้อสอบ (within-item multidimensional model) โมเดลทั้งสามประเภทนี้จะถูกตรวจสอบประสิทธิภาพในการบรรยายความสามารถนักเรียนและเปรียบเทียบความสอดคล้องกลมกลืนกับข้อมูลคำตอบข้อสอบชุดเดียวกันเพื่อหาโมเดลที่สอดคล้องกลมกลืนมากที่สุด (best fit model)

สำหรับโมเดลเอกมิติหรือโมเดล UniFit-50 ซึ่งเป็นโมเดลฐาน (baseline model) ค่าความแปรปรวนความสามารถนักเรียนเท่ากับ .57 และค่าความเที่ยงแบบ EAP/PV (EAR/PV reliability) ซึ่งหมายถึงสัดส่วนความแปรปรวนที่อธิบายได้ด้วยโมเดล UniFit-50 ต่อความแปรปรวนรวมทั้งหมดเท่ากับ .84 แสดงให้เห็นว่าโมเดลนี้อธิบายความแปรปรวนได้สูง ในขณะที่ความแปรปรวนรวมทั้งหมดมีค่าต่ำ นั่นคือค่าประมาณความสามารถนักเรียนที่ได้จากโมเดลนี้คงเส้นคงวาและเมื่อพิจารณาค่าสถิติทดสอบความสอดคล้องกลมกลืนของข้อสอบรายข้อแบบถ่วงน้ำหนัก (weighted fit mean square statistics) หรือค่าสถิติ infit (infit statistics) พบว่ามีค่า MNSQ ระหว่าง .94 ถึง 1.09 (ดูตารางที่ B1.1 ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์ระหว่าง .75 ถึง 1.33 (Adams & Khoo, 1996) ดังนั้นจึงไม่มี

ข้อสอบข้อใดที่ให้ข้อมูลไม่สอดคล้องกลมกลืน (misfit) กับข้อสอบข้ออื่นที่อยู่ในโมเดลเดียวกันและเป็นหลักฐานบ่งชี้ว่าผู้สร้างข้อสอบ ONET ปีการศึกษา 2550 มีแนวคิดชัดเจนในการออกแบบการวัดความสามารถทางวิทยาศาสตร์แบบองค์รวม

ตารางที่ 4.1 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2550

สาระการเรียนรู้	โมเดลพหุมิติระหว่างข้อสอบ		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8	
	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.60	.66	.30	.15	.65	.67
2: ชีวิตกับสิ่งแวดล้อม	1.61	.63	.63	.22	1.45	.59
3: สารและสมบัติของสาร	.73	.75	.23	.10	.73	.76
4: แรงและการเคลื่อนที่	1.41	.67	.58	.26	1.46	.69
5: พลังงาน	.67	.71	.27	.10	.66	.70
6: กระบวนการเปลี่ยนแปลงของโลก	.76	.69	.31	.11	.74	.68
7: ดาราศาสตร์และอวกาศ	1.03	.71	.35	.18	1.02	.55
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-	-	.68	.71	.26	.17

ผลวิเคราะห์จากโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-50 ในตารางที่ 4.1 แสดงค่าความแปรปรวนระหว่าง .60 ถึง 1.61 โดยสาระการเรียนรู้ที่ 1, 3, 5 และ 6 มีค่าความแปรปรวนใกล้เคียงกันซึ่งแสดงว่าความสามารถนักเรียนของแต่ละสาระการเรียนรู้ไม่แตกต่างกันมากเมื่อเปรียบเทียบกับความสามารถของนักเรียนในสาระการเรียนรู้ที่ 2 และเมื่อพิจารณาค่าความเที่ยงซึ่งอยู่ระหว่าง .66 ถึง .75 แสดงให้เห็นว่าโมเดล BimFit-50 สามารถอธิบายความแปรปรวนของความสามารถนักเรียนได้ในระดับปานกลางและมีความแปรปรวนอีกส่วนหนึ่งที่ไม่สามารถอธิบายได้ด้วยโมเดลนี้ ความสัมพันธ์ของค่าประมาณความสามารถของแต่ละสาระการเรียนรู้ในตารางที่ 4.2 ซึ่งมีค่าระหว่าง .67 ถึง .80 แสดงให้เห็นว่าในขณะที่ความสามารถนักเรียนในแต่ละสาระการเรียนรู้สัมพันธ์กันค่อนข้างสูงก็ยังคงมีความแตกต่างเกิดขึ้นเพียงพอที่จะบ่งชี้ว่าธรรมชาติเนื้อหาของแต่ละสาระการเรียนรู้ไม่เหมือนกัน นอกจากนี้เมื่อพิจารณาความสอดคล้องกลมกลืนของข้อสอบแบบ infit พบว่ามีค่า MNSQ ระหว่าง .94 ถึง 1.13 (ดูตารางที่ B1.2 ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์ที่กำหนดจึงสรุปได้ว่าไม่มีข้อสอบข้อใดสอดคล้องกลมกลืนคลาดเคลื่อนกับข้อสอบข้ออื่นๆ นอกจากนี้ค่า MNSQ ยังแสดงให้เห็นว่ากระบวนการสร้างข้อสอบ ONET ปีการศึกษา 2550 ที่กำหนดให้ข้อสอบแต่ละข้อวัดสาระการเรียนรู้แตกต่างกันเป็นมุมมองที่นำมาใช้ออกแบบการวัดความสามารถนักเรียนได้

ตารางที่ 4.2 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2550

สาระการเรียนรู้	สัมประสิทธิ์สหสัมพันธ์						
	1	2	3	4	5	6	7
โมเดลพหุมิติระหว่างข้อสอบ							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.68						
3: สารและสมบัติของสาร	.75	.76					
4: แรงและการเคลื่อนที่	.67	.71	.76				
5: พลังงาน	.72	.73	.80	.74			
6: กระบวนการเปลี่ยนแปลงของโลก	.71	.70	.78	.73	.77		
7: ดาราศาสตร์และอวกาศ	.73	.73	.79	.73	.78	.80	
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบทุกข้อวัดสาระที่ 8							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.05						
3: สารและสมบัติของสาร	.39	.23					
4: แรงและการเคลื่อนที่	.02	.25	.25				
5: พลังงาน	.34	.11	.45	.23			
6: กระบวนการเปลี่ยนแปลงของโลก	.30	.12	.39	.19	.39		
7: ดาราศาสตร์และอวกาศ	.16	.15	.24	.20	.29	.33	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-.41	.21	-.23	.20	-.34	-.25	.02
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบบางข้อวัดสาระที่ 8							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.63						
3: สารและสมบัติของสาร	.75	.73					
4: แรงและการเคลื่อนที่	.69	.70	.77				
5: พลังงาน	.72	.68	.78	.75			
6: กระบวนการเปลี่ยนแปลงของโลก	.69	.66	.76	.73	.74		
7: ดาราศาสตร์และอวกาศ	.61	.61	.68	.63	.66	.71	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	.13	.02	.10	.15	.09	.00	-.31

ผลวิเคราะห์จากโมเดลพหุมิติภายในข้อสอบที่กำหนดให้ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-allSt8-50) และที่กำหนดให้ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-someSt8-50) ในตารางที่ 4.1 พบว่า มีข้อค้นพบที่น่าสนใจหลายประการ กล่าวคือ เมื่อกำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 ในโมเดล WimFit-allSt8-50 ค่าความแปรปรวนความสามารถนักเรียนของแต่ละสาระการเรียนรู้ต่ำกว่าโมเดล BimFit-50 อย่างเห็นได้ชัด และเมื่อกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ตามความเห็นของผู้เชี่ยวชาญในโมเดล WimFit-someSt8-50 พบว่ามีค่าความแปรปรวนเพิ่มขึ้นและใกล้เคียงกับโมเดล BimFit-50 นอกจากนี้ความเที่ยงของสาระที่ 1 ถึง 7 ของ

โมเดล WimFit-allSt8-50 มีค่าต่ำกว่าโมเดล BimFit-50 และโมเดล WimFit-someSt8-50 นั้นหมายความว่าโมเดล WimFit-allSt8-50 อธิบายความแปรปรวนของความสามารถนักเรียนได้ดีและเป็นหลักฐานบ่งชี้ปัญหาการกำหนดลักษณะโมเดลคลาดเคลื่อน (model misspecification) หรือข้อสอบทุกข้อไม่ได้วัดสาระที่ 8 และในขณะที่โมเดล WimFit-someSt8-50 ซึ่งผู้เชี่ยวชาญได้ประเมินว่ามีข้อสอบจำนวน 5 ข้อวัดสาระที่ 8 (ดูตารางที่ 3.5 ในบทที่ 3) ก็ไม่สามารถอธิบายความแปรปรวนของความสามารถนักเรียนได้เพิ่มขึ้น แม้ว่าโมเดล WimFit-someSt8-50 จะรักษาค่าความเที่ยงของสาระที่ 1 ถึง 7 ให้ใกล้เคียงกับโมเดล BimFit-50 ก็ตาม

ความสัมพันธ์ของค่าประมาณความสามารถของแต่ละสาระการเรียนรู้จากโมเดล WimFit-allSt8-50 มีค่าต่ำและติดลบซึ่งแตกต่างอย่างชัดเจนกับโมเดล WimFit-someSt8-50 ที่มีค่าในระดับปานกลางยกเว้นสาระที่ 8 ซึ่งเมื่อพิจารณาค่าสหสัมพันธ์รวมกับค่าความแปรปรวนและความเที่ยงจึงพบว่าโมเดล WimFit-allSt8-50 ไม่เหมาะสมอย่างยิ่งต่อการประมาณค่าความสามารถนักเรียน แม้ว่าจะมีค่า MNSQ อยู่ในช่วงเกณฑ์ที่กำหนดก็ตาม นอกจากนี้แม้ว่าโมเดล WimFit-someSt8-50 จะมีคุณสมบัติดีกว่าโมเดล WimFit-allSt8-50 แต่ก็มีค่าสหสัมพันธ์ของสาระที่ 8 ต่ำและติดลบด้วย โดยเฉพาะอย่างยิ่งค่าสหสัมพันธ์ระหว่างสาระที่ 6 และ 8 ที่เท่ากับศูนย์ ทั้งๆ ที่มีข้อสอบสองข้อวัดทั้งสาระที่ 6 และ 8 ตามความเห็นของผู้เชี่ยวชาญ (ดูตารางที่ 3.5 ในบทที่ 3) ดังนั้นจึงไม่อาจกล่าวได้ว่าโมเดล WimFit-someSt8-50 เป็นโมเดลการวัดที่เหมาะสมที่สุดเช่นกัน

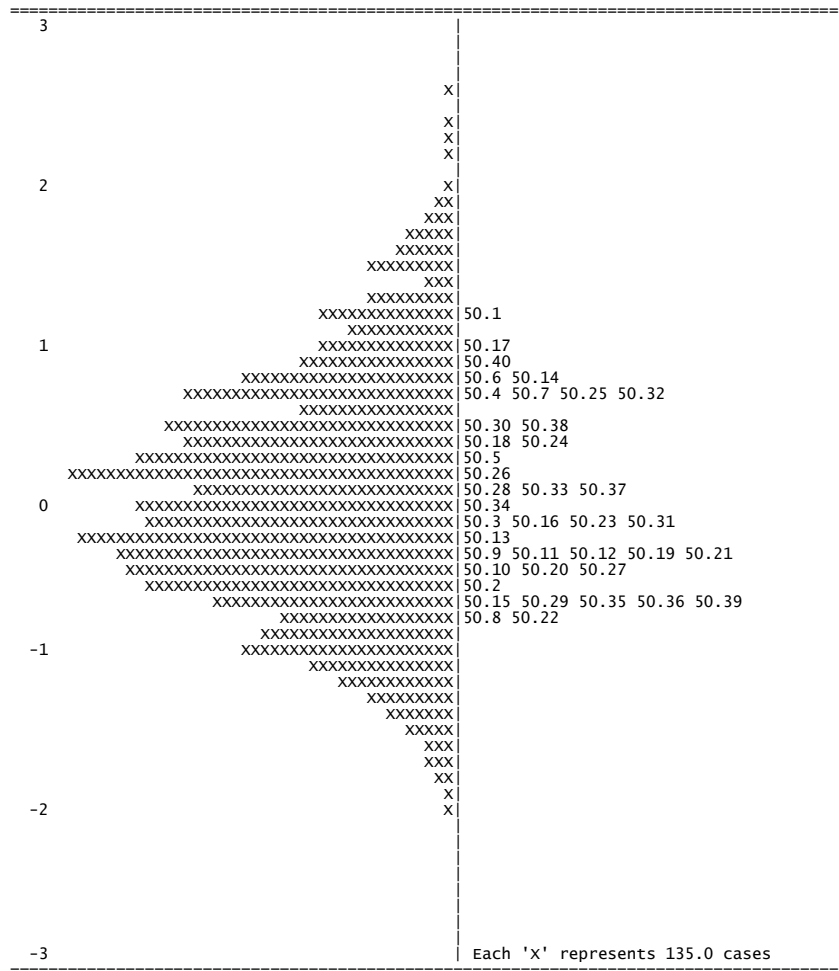
แผนที่ของ Wright ในภาพที่ 4.1(a) แสดงการแจกแจงค่าพารามิเตอร์ความสามารถนักเรียนและพารามิเตอร์ความยากข้อสอบที่ได้จากการประมาณด้วยโมเดลเอกมิติหรือโมเดล UniFit-50 โดยมีสเกลลอจิท (logit) ที่เป็นเส้นตรงแนวตั้งบอกความสัมพันธ์ของข้อมูลทั้งสองส่วน แผนที่ของ Wright มีประโยชน์อย่างมากในการตีความความสัมพันธ์ระหว่างตำแหน่งความสามารถนักเรียนและตำแหน่งความยากข้อสอบที่อยู่บนสเกลลอจิทเดียวกัน นักเรียนที่มีค่าประมาณความสามารถอยู่ในตำแหน่งเดียวกับความยากข้อสอบแสดงให้เห็นว่ามีโอกาส 50% ที่จะตอบข้อสอบข้อนั้นถูกต้อง ถ้าตำแหน่งค่าประมาณความสามารถนักเรียนสูงกว่าตำแหน่งความยากข้อสอบจะแสดงว่านักเรียนมีโอกาสตอบถูกข้อสอบข้อนั้นมากกว่า 50% และในทางตรงกันข้ามถ้าตำแหน่งความสามารถนักเรียนอยู่ต่ำกว่าตำแหน่งความยากข้อสอบก็จะแสดงว่านักเรียนมีโอกาสตอบถูกข้อสอบข้อนั้นน้อยกว่า 50% เมื่อพิจารณาค่าความยากข้อสอบพบว่ามีความห่าง -0.75 ถึง 1.12 ลอจิทซึ่งแคบกว่าค่าความสามารถนักเรียนที่มีความห่าง -1.78 ถึง 2.27 ลอจิทและมีข้อที่ 1 ยากที่สุดและข้อที่ 8 และ 22 ง่ายที่สุด ตำแหน่งความยากข้อสอบส่วนใหญ่กระจุกตัวอยู่บริเวณค่าลอจิทเท่ากับ 0 ซึ่งสอดคล้องกับตำแหน่งความสามารถนักเรียนส่วนใหญ่ด้วย นั่นคือข้อสอบส่วนใหญ่ยากปานกลางเมื่อเปรียบเทียบกับการแจกแจงค่าประมาณความสามารถนักเรียนและข้อสอบส่วนใหญ่สามารถใช้บรรยายความสามารถนักเรียนในระดับปานกลางได้ดีด้วย แต่อย่างไรก็ตามสำหรับนักเรียนมีค่าความสามารถต่ำกว่า -0.75 ลอจิทพบว่าไม่มีข้อสอบข้อใดเลยที่ง่ายพอเหมาะกับนักเรียนกลุ่มนี้ นั่นคือข้อสอบ ONET ยังไม่สามารถบรรยายความสามารถของนักเรียนในกลุ่มนี้ได้ และในทำนองเดียวกัน กลุ่มนักเรียนที่มีค่า

ความสามารถมากกว่า 1.12 ลอจิกก็ไม่มีข้อสอบข้อใดที่ยากเหมาะสมกับความสามารถของนักเรียนกลุ่มนี้

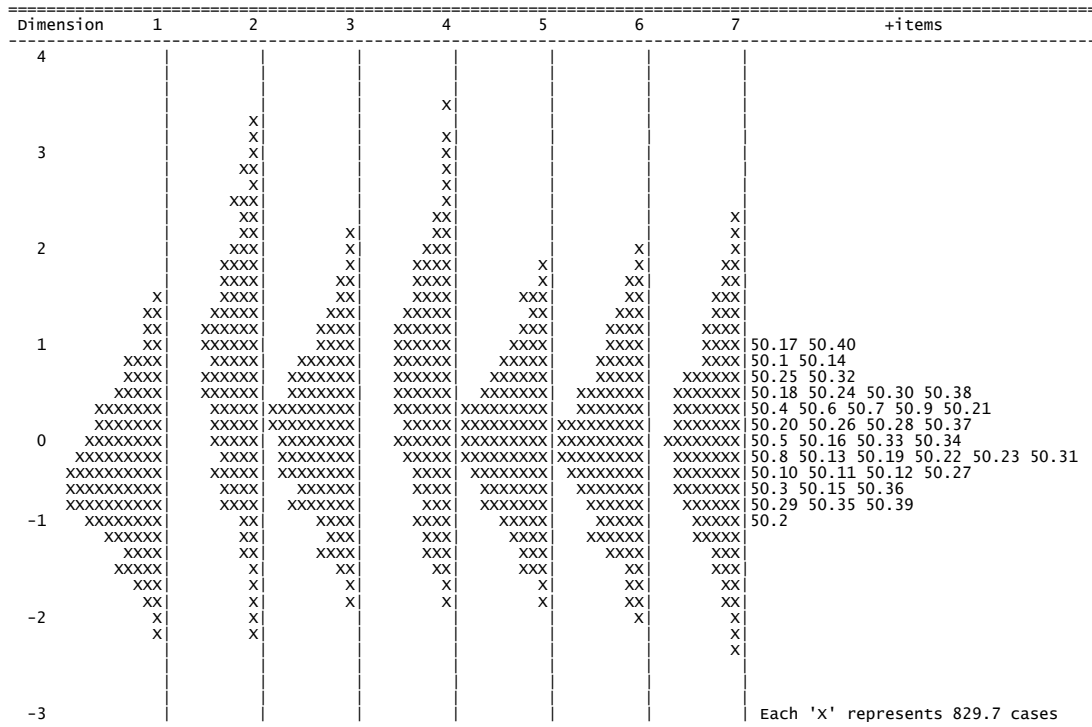
เมื่อพิจารณาแผนที่ของ Wright สำหรับโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-50 ในภาพที่ 4.1(b) พบว่า ตำแหน่งค่าความยากข้อสอบทั้งหมดอยู่บริเวณตรงกลางของการแจกแจงความสามารถนักเรียนในสาระที่ 3, 5, 6 และ 7 อยู่ต่ำกว่าเล็กน้อยในสาระที่ 2 และ 4 และอยู่สูงกว่าในสาระที่ 1 ซึ่งแสดงให้เห็นว่าในภาพรวมนักเรียนมีความสามารถในการเรียนรู้ไม่เท่ากัน ข้อสอบที่วัดสาระที่ 1 จึงเป็นข้อสอบยาก ส่วนข้อสอบที่เหลือในภาพรวมเป็นข้อสอบยากปานกลาง นอกจากนี้เมื่อพิจารณาการแจกแจงค่าความสามารถนักเรียนของแต่ละสาระการเรียนรู้พบว่ามีความแตกต่างกันอย่างชัดเจนโดยการแจกแจงค่าความสามารถในสาระที่ 2, 4 และ 7 มีลักษณะแบนแสดงว่านักเรียนมีความสามารถหลากหลายมากกว่าสาระที่ 3, 5 และ 6 ซึ่งมีลักษณะการแจกแจงที่โด่งกว่า เมื่อพิจารณาค่าความยากข้อสอบของโมเดล BimFit-50 ซึ่งอยู่ระหว่าง -.94 ถึง 1.03 ลอจิกพบว่าช่วงแคบกว่าโมเดล UniFit-50 และมีข้อที่ 17 เป็นข้อที่ยากที่สุด ในขณะที่มีข้อที่ 2 เป็นข้อที่ง่ายที่สุด

แผนที่ของ Wright ในภาพที่ 4.1(c) แสดงตำแหน่งความยากข้อสอบและตำแหน่งความสามารถของนักเรียนจากโมเดลที่กำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 หรือโมเดล WimFit-allSt8-50 ซึ่งค่าความยากข้อสอบอยู่ในช่วงแคบมากขึ้นระหว่าง -.45 ถึง .55 ลอจิกโดยมีข้อที่ 8 เป็นข้อง่ายที่สุดและมีข้อที่ 1 เป็นข้อยากที่สุดและแสดงให้เห็นว่าข้อสอบทั้งหมดยากปานกลาง ส่วนภาพที่ 4.1(d) เป็นแผนที่ของ Wright ที่ได้จากโมเดล WimFit-someSt8-50 ซึ่งกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ค่าความยากข้อสอบที่ได้จากโมเดลนี้อยู่ระหว่าง -.77 ถึง 1.12 ลอจิกซึ่งใกล้เคียงกับโมเดล BimFit-50 และมีข้อที่ 1 เป็นข้อยากที่สุดและข้อ 15, 22, 29, และ 35 เป็นกลุ่มข้อง่ายที่สุด ข้อสังเกตระหว่างแผนที่ Wright จากโมเดล WimFit-allSt8-50 และ WimFit-someSt8-50 คือเมื่อกำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 พบว่าการแจกแจงค่าความสามารถนักเรียนมีลักษณะแบนและมีช่วงค่าประมาณความสามารถกว้างแสดงว่าการกำหนดให้ข้อสอบทั้งหมด 40 ข้อวัดสาระที่ 8 นั้นจำแนกความสามารถนักเรียนได้ด้นน้อยกว่าข้อสอบเพียงห้าข้อที่กำหนดไว้ในโมเดล WimFit-someSt8-50 นอกจากนี้การแจกแจงค่าความสามารถในสาระที่ 2, 4 และ 7 ของโมเดล WimFit-allSt8-50 และโมเดล WimFit-someSt8-50 ก็มีลักษณะแบนเช่นเดียวกับโมเดล BimFit-50 ซึ่งสะท้อนว่าสาระการเรียนรู้ทั้งสามนี้ต้องการข้อสอบเพิ่มขึ้นเพื่อบรรยายความสามารถที่แตกต่างกันในแต่ละสาระการเรียนรู้

นอกจากนี้แผนที่ของ Wright ที่ได้จากโมเดลพหุมิติระหว่างข้อสอบและภายในข้อสอบยังแสดงสารสนเทศสำคัญเช่นเดียวกับโมเดลเอกมิติที่ว่าแบบสอบ ONET ปีการศึกษา 2550 ขาดข้อสอบที่มียากเหมาะสมกับนักเรียนที่มีความสามารถสูง และขาดข้อสอบที่ง่ายเหมาะสมกับนักเรียนที่มีความสามารถต่ำ เพราะระดับความยากข้อสอบที่ต่างกันจะเป็นเครื่องมือสำหรับพัฒนาโครงสร้างความสามารถนักเรียนเพื่อใช้บรรยายความรู้และทักษะที่นักเรียนแต่ละคนมีไม่เหมือนกัน



(a) โมเดลเอกมิติ



(b) โมเดลพหุมิติระหว่างข้อสอบ

Dimension	1	2	3	4	5	6	7	8	+items
2									
		X						X	
		X						X	
		X						XX	
		X						X	
1	X	XX	X	XX	X	X	X	XX	
	X	XXX	X	XX	X	XX	XX	XX	
	XX	XX	XX	XX	X	X	XX	XX	
	XX	XXXX	XX	XXXX	XX	XXXX	XXXX	XXXX	
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.1
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.17 50.40
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.6 50.7 50.14 50.25
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.4 50.30 50.32 50.38
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.5 50.18 50.24
0	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.26 50.28 50.33 50.37
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.3 50.16 50.23 50.31 50.34
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.11 50.12 50.13 50.19 50.21 50.27
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.2 50.9 50.10 50.20 50.35 50.36
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.15 50.22 50.29 50.39
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.8
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
	XX	XX	XX	XX	XX	XX	XX	XX	
-1	X	XX	X	XX	X	X	X	XX	
	X	XX	X	XX	X	X	X	XX	
	X	XX	X	XX	X	X	X	XX	
	X	XX	X	XX	X	X	X	XX	
-2		X							

Each 'x' represents 911.9 cases

(c) โมเดลพหุมิติภายในข้อสอบโดยที่ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8

Dimension	1	2	3	4	5	6	7	8	+items
3									
		X		X					
		X		X					
		X		X					
2		X		X			X		
		XX	X	X		X	X		
	X	X	X	XX	X	X	X		
	XX	XX	XX	XX	XX	XX	XX		
1	XX	XX	XX	XX	XX	XX	XX	X	
	XX	XX	XX	XX	XX	XX	XX	XX	50.1
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XX	50.17
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XX	50.6 50.7 50.14 50.25 50.32
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.4
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.18 50.24 50.30 50.40
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.5 50.38
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.26 50.28 50.33 50.37
0	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.16 50.23 50.31 50.34
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.3 50.10 50.13
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.8 50.9 50.11 50.12 50.19 50.20 50.21 50.27 50.36 50.39
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.2
	XXXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	50.15 50.22 50.29 50.35
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
	XX	XX	XX	XX	XX	XX	XX	XX	
-1	X	XX	X	XX	X	X	X	XX	
	X	XX	X	XX	X	X	X	XX	
	X	XX	X	XX	X	X	X	XX	
-2		X	X	X		X	X		
	X	X	X	X		X	X		
-3		X							

Each 'x' represents _BIG_ cases

(d) โมเดลพหุมิติภายในข้อสอบโดยที่ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8

ภาพที่ 4.1 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2550

ตารางที่ 4.3 แสดงโมเดลการวัดสำหรับเปรียบเทียบความสอดคล้องกลมกลืน โดยแต่ละโมเดลมีข้อกำหนดลักษณะโมเดลแตกต่างกันจึงทำให้มีจำนวนพารามิเตอร์ที่ต้องการประมาณด้วยกรอบโมเดล MRCML ไม่เท่ากัน กล่าวคือ โมเดลเอกมิติหรือโมเดล UniFit-50 มีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากับ 41 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 39 ค่า (เนื่องจากกรอบโมเดล MRCML จะกำหนดให้ผลรวมของค่าพารามิเตอร์ความยากทั้งหมดที่ประมาณได้จากโมเดลมีค่าเท่ากับศูนย์) ค่าเฉลี่ยความสามารถของประชากรนักเรียนจำนวนหนึ่งค่า และความ

แปรปรวนความสามารถของประชากรนักเรียนจำนวนหนึ่งค่า สำหรับโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-50 จะมีพารามิเตอร์ที่ต้องประมาณจำนวน 68 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 33 ค่า (ผลรวมของค่าพารามิเตอร์ความยากของแต่ละสาระการเรียนรู้เท่ากับศูนย์) ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนอย่างละจำนวน 7 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 21 ค่า และสำหรับโมเดลพหุมิติภายในข้อสอบทั้งโมเดล WimFit-allSt8-50 และโมเดล WimFit-someSt8-50 จะมีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากันคือ 76 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 32 ค่า ค่าเฉลี่ยและความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 8 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 28 ค่า

ตารางที่ 4.3 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2550

โมเดล	จำนวนพารามิเตอร์	AIC	BIC	G ²
โมเดลเอกมิติ	41	4545525.44	4545646.56	4545443.44
โมเดลพหุมิติระหว่างข้อสอบ	68	4542927.32	4543128.21	4542791.32
โมเดลพหุมิติภายในข้อสอบโดย ข้อสอบทุกข้อวัดสาระที่ 8	76	4543928.88	4544153.40	4543776.88
โมเดลพหุมิติภายในข้อสอบโดย ข้อสอบบางข้อวัดสาระที่ 8	76	4545113.49	4545338.02	4544961.49

เนื่องจากโมเดลพหุมิติมีความสัมพันธ์แบบระดับลดหลั่นกับโมเดลเอกมิติ ดังนั้นการเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลจึงพิจารณาจากหลายวิธี ไม่ว่าจะเป็นผลต่างของค่า deviance (G²) โดยผลต่างของ G² ระหว่างโมเดลลดรูป (reduced model) หรือโมเดลฐาน (baseline model) กับโมเดลแข่งขัน (competing model) ที่สนใจจะมีการแจกแจงใกล้เคียงกับการแจกแจงแบบ chi-square ที่มี degree of freedom (df) เท่ากับผลต่างของ df จากทั้งสองโมเดล (Briggs & Wilson, 2003) และสามารถตรวจสอบได้จากการเปรียบเทียบค่า G², AIC และ BIC ระหว่างโมเดลต่างๆ โดยโมเดลที่มีค่า G², AIC และ BIC น้อยที่สุดถือว่าเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลคำตอบข้อสอบมากที่สุด เมื่อพิจารณาผลวิเคราะห์ในตารางที่ 4.3 พบว่าโมเดลพหุมิติทั้งสามโมเดลสอดคล้องกลมกลืนกับข้อมูลดีกว่าโมเดลเอกมิติอย่างมีนัยสำคัญทางสถิติ (โมเดล Unifit-50 กับ BimFit-50: $\chi^2=2,652.12$, $df=27$, $p<.01$; โมเดล UniFit-50 กับ WimFit-allSt8-50: $\chi^2=1,666.56$, $df=35$, $p<.01$; โมเดล UniFit-50 กับ WimFit-someSt8-50: $\chi^2=481.95$, $df=35$, $p<.01$) และเมื่อพิจารณาโมเดลพหุมิติทั้งสามโมเดลพบว่าโมเดล BimFit-50 มีค่า G², AIC และ BIC น้อยที่สุดโดยสอดคล้องกลมกลืนกับข้อมูลดีกว่าโมเดล WimFit-allSt8-50 ($\chi^2=985.56$, $df=8$, $p<.01$) และดีกว่าโมเดล WimFit-someSt8-50 ($\chi^2=2,170.17$, $df=8$, $p<.01$) ดังนั้นจึงสรุปว่า

โมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-50 เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด แสดงว่าจำนวนพารามิเตอร์ต่างๆ ที่เพิ่มเข้าไปในโมเดลตามข้อกำหนดลักษณะแบบสอบและโครงสร้างความสามารถทางวิทยาศาสตร์ของนักเรียนตามหลักสูตรแกนกลางฯ พ.ศ. 2551 ให้สารสนเทศที่เป็นประโยชน์ต่อการอธิบายความสามารถทางวิทยาศาสตร์ของนักเรียนได้เพิ่มขึ้น

ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2551

โมเดลเอกมิติหรือโมเดล UniFit-51 เป็นโมเดลฐานของการวัดความสามารถนักเรียนด้วยข้อสอบ ONET ปีการศึกษา 2551 ค่าความแปรปรวนความสามารถนักเรียนเท่ากับ .57 และค่าความเที่ยงแบบ EAP/PV เท่ากับ .83 แสดงให้เห็นว่าความแปรปรวนที่อธิบายได้ด้วยโมเดลนี้มีค่าสูง นั่นคือค่าประมาณความสามารถนักเรียนที่ได้จากโมเดลนี้คงเส้นคงวาและเมื่อพิจารณาค่าสถิติทดสอบความสอดคล้องกลมกลืนของข้อสอบรายข้อแบบถ่วงน้ำหนัก พบว่ามีค่า MNSQ ระหว่าง .86 ถึง 1.11 (ดูตารางที่ B2.1 ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์ขนาดอิทธิพล (effect size) ที่ยอมรับ ดังนั้นจึงไม่มีข้อสอบข้อใดแสดงความสอดคล้องกลมกลืนคลาดเคลื่อนไปจากกับข้อสอบข้ออื่นที่อยู่ในโมเดล UniFit-51 และเป็นหลักฐานบ่งชี้ว่ามุมมองการวัดความสามารถทางวิทยาศาสตร์แบบองค์รวมสามารถนำมาใช้สร้างแบบสอบ ONET ปีการศึกษา 2551 ได้

ตารางที่ 4.4 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2551

สาระการเรียนรู้	โมเดลพหุมิติระหว่างข้อสอบ		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8	
	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.71	.68	.33	.11	.70	.70
2: ชีวิตกับสิ่งแวดล้อม	1.02	.62	.48	.15	.97	.61
3: สารและสมบัติของสาร	.98	.75	.36	.20	1.00	.74
4: แรงและการเคลื่อนที่	1.17	.61	.49	.15	1.09	.62
5: พลังงาน	.77	.68	.32	.11	.79	.67
6: กระบวนการเปลี่ยนแปลงของโลก	.77	.69	.30	.09	.74	.66
7: ดาราศาสตร์และอวกาศ	.76	.68	.30	.12	.75	.70
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-	-	.66	.65	.20	.17

ผลวิเคราะห์จากโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-51 ในตารางที่ 4.4 แสดงค่าความแปรปรวนระหว่าง .71 ถึง 1.17 โดยสาระการเรียนรู้ที่ 1, 5, 6 และ 7 มีค่าความแปรปรวน

ใกล้เคียงกันซึ่งแสดงให้เห็นว่าความสามารถนักเรียนในสาระการเรียนรู้เหล่านี้ไม่แตกต่างกันมากเมื่อเปรียบเทียบกับสาระการเรียนรู้ที่ 4 และเมื่อพิจารณาค่าความเที่ยงซึ่งอยู่ระหว่าง .61 ถึง .75 แสดงให้เห็นว่าโมเดล BimFit-50 สามารถอธิบายความแปรปรวนในความสามารถนักเรียนระดับปานกลาง ความสัมพันธ์ของค่าประมาณความสามารถนักเรียนแต่ละสาระการเรียนรู้ในตารางที่ 4.5 ซึ่งมีค่าระหว่าง .63 ถึง .76 แสดงให้เห็นว่ามีความแตกต่างเกิดขึ้นเพียงพอที่จะบ่งชี้ว่าธรรมชาติเนื้อหาของแต่ละสาระการเรียนรู้ที่วัดด้วยข้อสอบ ONET แต่ละข้อไม่เหมือนกัน นอกจากนี้เมื่อพิจารณาความสอดคล้องกลมกลืนของข้อสอบแบบ infit พบว่ามีค่า MNSQ ระหว่าง .90 ถึง 1.12 (ดูตารางที่ B2.2 ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์นัยสำคัญทางปฏิบัติ (practical significance) จึงสรุปได้ว่าไม่มีหลักฐานบ่งชี้ว่าข้อสอบสอดคล้องกลมกลืนคลาดเคลื่อนกับข้อสอบข้ออื่น นอกจากนี้ค่า MNSQ ที่อยู่ในช่วงเกณฑ์ดังกล่าวยังสามารถตีความได้ว่าโค้งคุณลักษณะข้อสอบเชิงประจักษ์ (empirical item characteristic curve) ที่สร้างจากคำตอบนักเรียนสามารถตรวจจับ (capture) ลักษณะโค้งคุณลักษณะข้อสอบที่สร้างขึ้นจากค่าประมาณความสามารถนักเรียน (estimated item characteristic curve) ได้ดี นั่นหมายความว่าข้อสอบ ONET ปีการศึกษา 2551 สามารถจำแนกความสามารถที่แตกต่างกันของนักเรียนได้ดีเช่นเดียวกับโมเดล UniFit-51

ผลวิเคราะห์จากโมเดลพหุมิติภายในข้อสอบทั้งประเภทที่กำหนดให้ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-allSt8-51) และที่กำหนดให้ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-someSt8-51) ในตารางที่ 4.4 พบว่า เมื่อกำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 ในโมเดล WimFit-allSt8-51 ค่าความแปรปรวนความสามารถนักเรียนของแต่ละสาระการเรียนรู้้น้อยกว่าโมเดล BimFit-51 อย่างชัดเจน และเมื่อกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ตามความเห็นของผู้เชี่ยวชาญในโมเดล WimFit-someSt8-51 พบว่าค่าความแปรปรวนใกล้เคียงกับโมเดล BimFit-51 นอกจากนี้ความเที่ยงของสาระที่ 1 ถึง 7 ของโมเดล WimFit-allSt8-51 มีค่าต่ำกว่าโมเดล BimFit-51 และโมเดล WimFit-someSt8-51 แสดงให้เห็นว่าโมเดล WimFit-allSt8-51 อธิบายความแปรปรวนของความสามารถนักเรียนได้ต่ำและบ่งชี้ปัญหาการกำหนดลักษณะโมเดลคลาดเคลื่อนหรือข้อสอบทุกข้อไม่ได้วัดสาระที่ 8 ในขณะที่โมเดล WimFit-someSt8-51 ซึ่งผู้เชี่ยวชาญได้ประเมินว่ามีข้อสอบจำนวน 9 ข้อวัดสาระที่ 8 (ดูตารางที่ 3.5 ในบทที่ 3) ก็ไม่สามารถอธิบายความแปรปรวนของความสามารถนักเรียนในสาระนี้เพิ่มขึ้น แม้ว่าโมเดล WimFit-someSt8-51 จะรักษาค่าความเที่ยงของสาระที่ 1 ถึง 7 ให้ใกล้เคียงกับโมเดล BimFit-51 ก็ตาม

ความสัมพันธ์ของค่าประมาณความสามารถในแต่ละสาระการเรียนรู้จากโมเดล WimFit-allSt8-51 มีค่าต่ำและติดลบโดยเฉพาะอย่างยิ่งสาระที่ 8 ซึ่งแตกต่างอย่างชัดเจนกับโมเดล WimFit-someSt8-51 ที่มีค่าในระดับปานกลางยกเว้นสาระที่ 8 เมื่อพิจารณาค่าสหสัมพันธ์รวมกับค่าความแปรปรวนและความเที่ยงพบว่าโมเดล WimFit-allSt8-51 ไม่เหมาะสมที่จะใช้ประมาณความสามารถนักเรียนแม้ว่าค่า MNSQ ซึ่งอยู่ระหว่าง .89 ถึง 1.13 จะอยู่ในช่วงเกณฑ์ที่กำหนดก็ตาม นอกจากนี้แม้ว่าโมเดล WimFit-someSt8-51 จะมีค่าความเที่ยงสูงกว่าโมเดล WimFit-allSt8-51 แต่

ก็มีค่าสหสัมพันธ์ของสาระที่ 8 ต่ำและติดลบด้วย ดังนั้นจึงไม่อาจกล่าวได้ว่าโมเดล WimFit-someSt8-51 เป็นโมเดลการวัดที่เหมาะสมที่สุดเช่นกัน

ตารางที่ 4.5 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2551

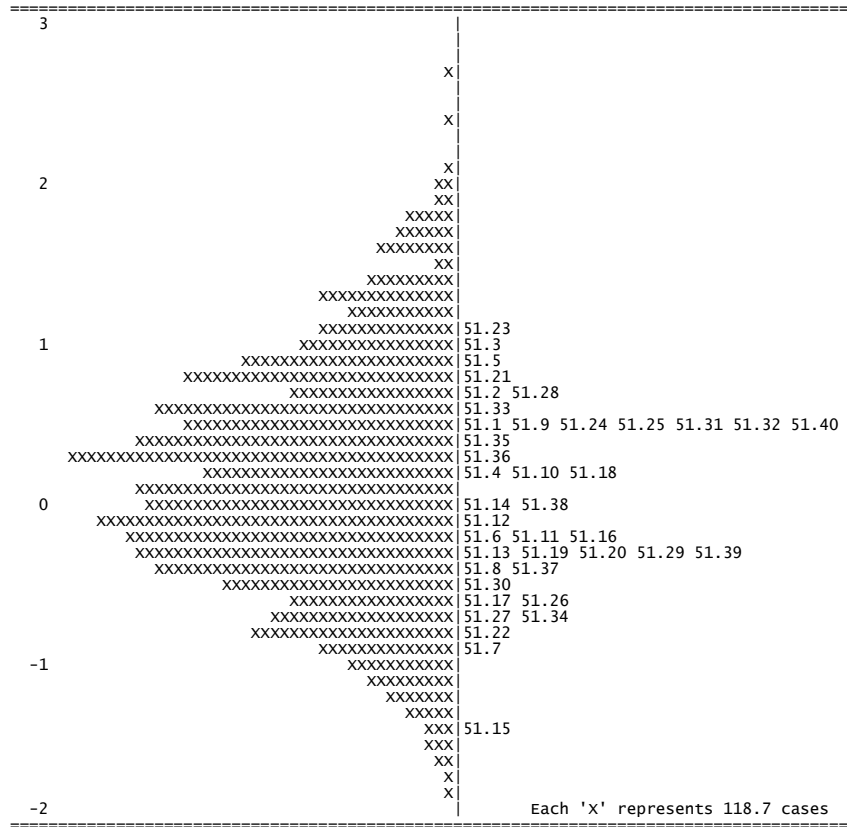
สาระการเรียนรู้	สัมประสิทธิ์สหสัมพันธ์						
	1	2	3	4	5	6	7
โมเดลพหุมิติระหว่างข้อสอบ							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.70						
3: สารและสมบัติของสาร	.75	.72					
4: แรงและการเคลื่อนที่	.67	.63	.70				
5: พลังงาน	.71	.66	.75	.68			
6: กระบวนการเปลี่ยนแปลงของโลก	.72	.68	.76	.67	.72		
7: ดาราศาสตร์และอวกาศ	.70	.67	.75	.68	.72	.73	
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบทุกข้อวัดสาระที่ 8							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.33						
3: สารและสมบัติของสาร	.38	.34					
4: แรงและการเคลื่อนที่	.22	.17	.28				
5: พลังงาน	.35	.23	.38	.31			
6: กระบวนการเปลี่ยนแปลงของโลก	.38	.27	.41	.20	.35		
7: ดาราศาสตร์และอวกาศ	.35	.24	.36	.22	.36	.41	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-.29	-.05	-.04	-.03	-.22	-.28	-.24
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบบางข้อวัดสาระที่ 8							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.73						
3: สารและสมบัติของสาร	.78	.72					
4: แรงและการเคลื่อนที่	.70	.64	.72				
5: พลังงาน	.73	.66	.77	.70			
6: กระบวนการเปลี่ยนแปลงของโลก	.74	.66	.79	.68	.75		
7: ดาราศาสตร์และอวกาศ	.74	.71	.77	.70	.73	.75	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	.01	.06	-.07	-.03	-.13	-.18	.06

แผนที่ของ Wright ในภาพที่ 4.2(a) แสดงการแจกแจงค่าความสามารถนักเรียน (ด้านซ้ายของภาพ) และค่าความยากข้อสอบ (ด้านขวาของภาพ) ที่ได้จากการประมาณด้วยโมเดลเอกมิติหรือโมเดล UniFit-51 เมื่อพิจารณาค่าความยากข้อสอบพบว่ามีค่าระหว่าง -1.45 ถึง 1.11 ลอจิทซึ่งแคบกว่าค่าความสามารถนักเรียนที่มีค่าระหว่าง -1.76 ถึง 2.27 ลอจิทและมีข้อที่ 23 ยากที่สุดและข้อที่ 1

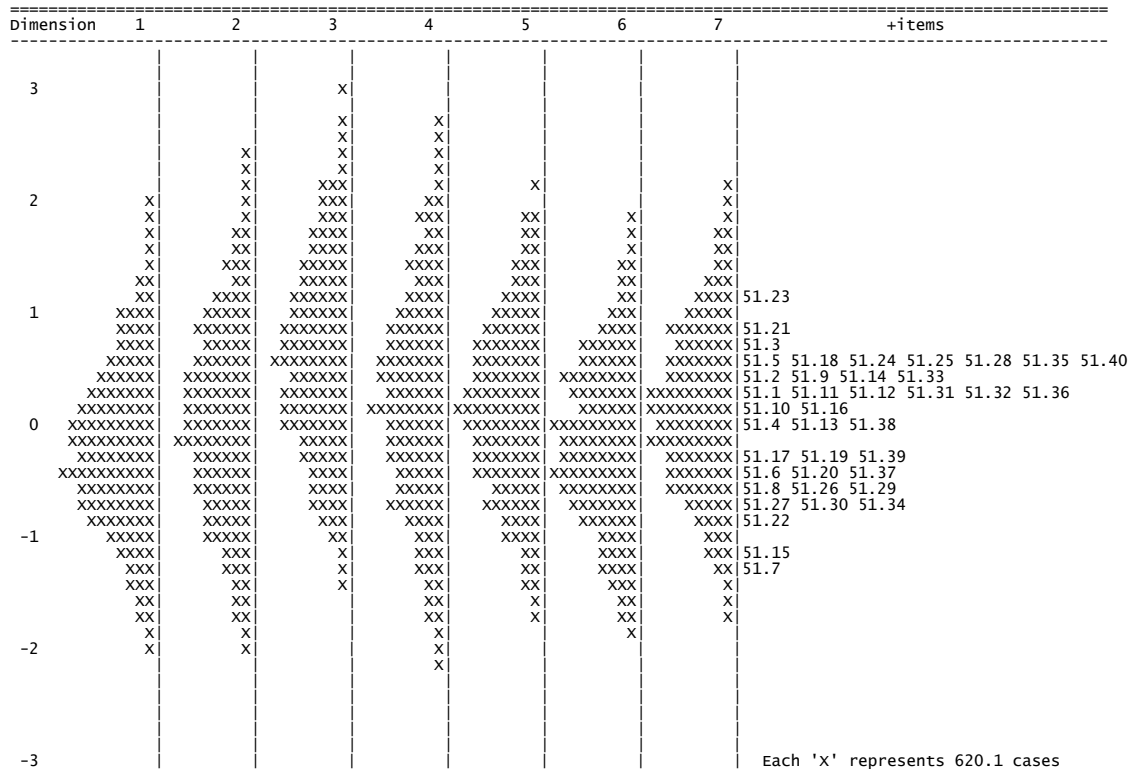
ง่ายที่สุด ตำแหน่งความยากข้อสอบมีการกระจายที่น่าสนใจคือ ข้อสอบครึ่งหนึ่งจะมีตำแหน่งสูงกว่า ลอจิตที่ศูนย์และข้อสอบทั้งหมดยกเว้นข้อที่ 15 มีตำแหน่งค่าความยากกระจุกตัวอยู่บริเวณตรงกลาง ของการแจกแจงค่าความสามารถนักเรียน ดังนั้นข้อสอบส่วนใหญ่ยากปานกลางเมื่อเปรียบเทียบกับ การแจกแจงค่าความสามารถนักเรียนและข้อสอบส่วนใหญ่สามารถใช้บรรยายความสามารถนักเรียน ในระดับปานกลางได้ดีด้วย อย่างไรก็ตามพบว่าไม่มีข้อสอบข้อใดเหมาะสมกับนักเรียนที่มีค่า ความสามารถสูงกว่า 1.11 ลอจิต นั่นคือข้อสอบ ONET ไม่สามารถบรรยายความสามารถนักเรียน กลุ่มนี้ได้

เมื่อพิจารณาแผนที่ของ Wright สำหรับโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-51 ในภาพที่ 4.2(b) พบว่า ตำแหน่งค่าความยากข้อสอบส่วนใหญ่อยู่บริเวณตรงกลางของการแจกแจง ความสามารถนักเรียนในทุกสาระการเรียนรู้ยกเว้นสาระที่ 3 ที่อยู่ต่ำกว่าแสดงให้เห็นว่าในภาพรวม นักเรียนส่วนใหญ่มีความสามารถในแต่ละสาระการเรียนรู้ใกล้เคียงกัน ข้อสอบที่วัดสาระที่ 3 มี แนวโน้มเป็นข้อสอบค่อนข้างง่ายเมื่อเปรียบเทียบกับข้อสอบสาระอื่นซึ่งอยู่ในระดับยากปานกลาง นอกจากนี้เมื่อพิจารณาการแจกแจงค่าความสามารถนักเรียนของแต่ละสาระการเรียนรู้พบว่าแตกต่างกันชัดเจนโดยการแจกแจงค่าความสามารถในสาระที่ 2, 3 และ 4 มีลักษณะแบน แสดงว่านักเรียนมี ความสามารถหลากหลายมากกว่าสาระที่ 1, 5, 6 และ 7 ลักษณะการแจกแจงที่แบนแสดงให้เห็นว่า ข้อสอบในสาระที่ 2, 3 และ 4 สามารถจำแนกความสามารถนักเรียนได้ดีน้อยกว่าข้อสอบในสาระอื่น เมื่อพิจารณาค่าความยากข้อสอบของโมเดล BimFit-51 ซึ่งอยู่ระหว่าง -1.31 ถึง 1.15 ลอจิตพบว่า ช่วงกว้างกว่าโมเดล UniFit-51 และมีข้อที่ 23 เป็นข้อยากที่สุด ในขณะที่ข้อที่ 7 เป็นข้อง่ายที่สุด

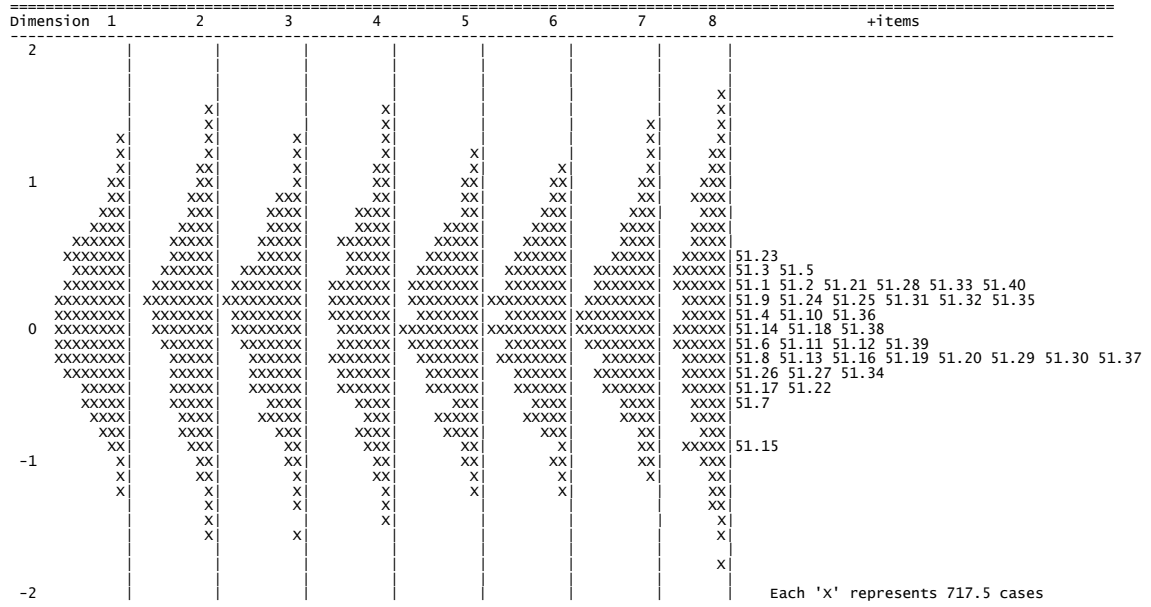
แผนที่ของ Wright ในภาพที่ 4.2(c) แสดงตำแหน่งค่าความยากข้อสอบและตำแหน่ง ความสามารถนักเรียนจากโมเดลที่กำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 หรือโมเดล WimFit-allSt8-51 ซึ่งค่าความยากข้อสอบอยู่ในช่วงแคบมากขึ้นระหว่าง -.87 ถึง .50 ลอจิต เมื่อเปรียบเทียบกับ ลักษณะการแจกแจงค่าความสามารถนักเรียน โดยมีข้อที่ 15 เป็นข้อง่ายที่สุดและมีข้อที่ 23 เป็นข้อ ยากที่สุดและแสดงให้เห็นว่าข้อสอบส่วนใหญ่เป็นยากปานกลางเมื่อเปรียบเทียบกับลักษณะการแจก แจงค่าความสามารถ ส่วนภาพที่ 4.2(d) เป็นแผนที่ของ Wright ที่ได้จากโมเดล WimFit-someSt8-51 ซึ่งกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ค่าความยากข้อสอบที่ได้จากโมเดลนี้อยู่ระหว่าง -1.74 ถึง 1.00 ลอจิตและมีข้อที่ 23 เป็นข้อยากที่สุดและข้อ 15 เป็นข้อง่ายที่สุด ข้อสังเกตระหว่างแผนที่ Wright จากโมเดล WimFit-allSt8-51 และ WimFit-someSt8-51 คือเมื่อกำหนดให้ข้อสอบทุกข้อ วัดสาระที่ 8 พบว่าการแจกแจงค่าความสามารถนักเรียนมีลักษณะแบนและมีช่วงค่าประมาณ ความสามารถกว้างแสดงว่าการกำหนดให้ข้อสอบทั้งหมด 40 ข้อวัดสาระที่ 8 นั้นจำแนกความสามารถ นักเรียนได้ดีน้อยกว่าข้อสอบ 9 ข้อที่กำหนดไว้ในโมเดล WimFit-someSt8-51 ตามความคิดเห็นของ ผู้เชี่ยวชาญ นอกจากนี้การแจกแจงค่าความสามารถในสาระที่ 1, 4 และ 8 ของโมเดล WimFit-allSt8-51 และการแจกแจงค่าความสามารถในสาระที่ 1 ถึง 7 ของโมเดล WimFit-someSt8-51 ก็มี ลักษณะแบนเมื่อเปรียบเทียบกับสาระอื่นในโมเดลเดียวกัน สะท้อนว่าโมเดลพหุมิติ



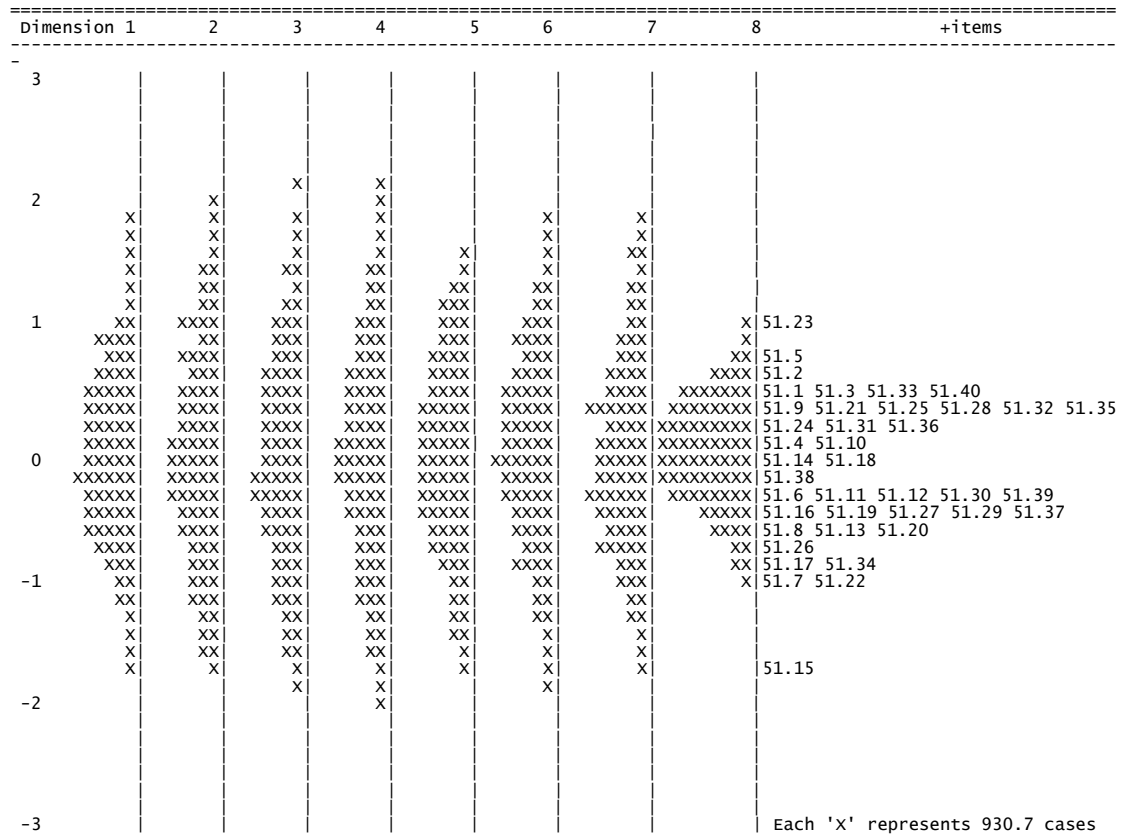
(a) โมเดลเอกมิติ



(b) โมเดลพหุมิติระหว่างข้อสอบ



(c) โมเดลพหุมิติภายในข้อสอบโดยที่ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8



(d) โมเดลพหุมิติภายในข้อสอบโดยที่ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8

ภาพที่ 4.2 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2551

ภายในข้อสอบทั้งสองโมเดลนี้ยังมีข้อจำกัดในการบรรยายความสามารถนักเรียน นอกจากนี้แผนที่ของ Wright ที่ได้จากโมเดลพหุมิติระหว่างข้อสอบและภายในข้อสอบยังแสดงสารสนเทศสำคัญ เช่นเดียวกับโมเดลเอกมิติที่ว่าแบบสอบ ONET ปีการศึกษา 2551 ยังขาดข้อสอบที่ยากเหมาะสมกับนักเรียนที่มีความสามารถสูง และขาดข้อสอบที่ง่ายเหมาะสมกับนักเรียนที่มีความสามารถต่ำด้วย

ตารางที่ 4.6 แสดงโมเดลพหุมิติจำนวนสามโมเดลสำหรับเปรียบเทียบความสอดคล้องกลมกลืนกับโมเดลฐานหรือโมเดลเอกมิติ โดยแต่ละโมเดลมีจำนวนพารามิเตอร์ที่ต้องการประมาณไม่เท่ากัน กล่าวคือโมเดลเอกมิติหรือโมเดล UniFit-51 มีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากับ 41 ค่า ได้แก่พารามิเตอร์ความยากข้อสอบจำนวน 39 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละหนึ่งค่า สำหรับโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-51 จะมีพารามิเตอร์ที่ต้องประมาณจำนวน 68 ค่า ได้แก่พารามิเตอร์ความยากข้อสอบจำนวน 33 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 7 ค่า และค่าความแปรปรวนร่วมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 21 ค่า และสำหรับโมเดลพหุมิติภายในข้อสอบทั้งโมเดล WimFit-allSt8-51 และโมเดล WimFit-someSt8-51 จะมีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากันคือ 76 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 32 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 8 ค่า และค่าความแปรปรวนร่วมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 28 ค่า

ตารางที่ 4.6 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2551

โมเดล	จำนวนพารามิเตอร์	AIC	BIC	G ²
โมเดลเอกมิติ	41	3894806.39	3894924.88	3894724.39
โมเดลพหุมิติระหว่างข้อสอบ	68	3897345.19	3897541.71	3897209.19
โมเดลพหุมิติภายในข้อสอบโดย ข้อสอบทุกข้อวัดสาระที่ 8	76	3896109.39	3896329.03	3895957.39
โมเดลพหุมิติภายในข้อสอบโดย ข้อสอบบางข้อวัดสาระที่ 8	76	3895184.26	3895403.90	3895032.26

เมื่อเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลในตารางที่ 4.6 พบว่าโมเดลเอกมิติสอดคล้องกลมกลืนกับข้อมูลดีกว่าโมเดลพหุมิติทั้งหมดเนื่องจากมีค่า G² น้อยที่สุดและผลต่างของค่า G² มีนัยสำคัญทางสถิติ (โมเดล Unifit-51 กับ BimFit-51: $\chi^2=2,484.80$, df=27, p<.01; โมเดล UniFit-51 กับ WimFit-allSt8-51: $\chi^2=1,233.00$, df=35, p<.01; โมเดล UniFit-51 กับ WimFit-someSt8-51: $\chi^2=307.87$, df=35, p<.01) และเมื่อพิจารณาค่า AIC และ BIC ของโมเดลเอกมิติพบว่ามีค่าน้อยที่สุดเช่นกัน แสดงว่าโมเดลเอกมิติเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด

ความสอดคล้องกลมกลืนของโมเดลกับคำตอบข้อสอบ ONET ปีการศึกษา 2552

ผลการวิจัยในส่วนนี้มีเนื้อหาแตกต่างจากปี 2550 และ 2551 เนื่องจากโครงการวิจัยนี้กำหนดให้ข้อสอบ ONET ปี 2552 มีวิธีให้คะแนนคำตอบสองแบบคือข้อสอบทุกข้อมีคะแนนสองค่า ซึ่งเป็นวิธีดั้งเดิมของ สทศ. และแบบหลายค่าเฉพาะข้อที่ 33 ถึง 37 โดยข้อดังกล่าวจะได้คะแนนความรู้บางส่วนเมื่อนักเรียนตอบถูกหนึ่งตัวเลือกจากทั้งหมดสองตัวเลือก ดังนั้นโมเดลการวัดความสามารถทางวิทยาศาสตร์ ปีการศึกษา 2552 จึงมีสองชุดคือรหัส D สำหรับข้อมูลที่มีสองค่าและรหัส P สำหรับข้อมูลที่มีหลายค่า

โมเดลเอกมิติแบบข้อมูลสองค่าหรือโมเดล UniFit-52D และแบบข้อมูลหลายค่าหรือโมเดล UniFit-52P เป็นโมเดลฐานของปีการศึกษา 2552 ค่าความแปรปรวนความสามารถนักเรียนของโมเดล UniFit-52D เท่ากับ .45 และค่าความเที่ยงแบบ EAP/PV เท่ากับ .77 และโมเดล UniFit-52P มีค่าความแปรปรวนเท่ากับ .34 และค่าความเที่ยงแบบ EAP/PV เท่ากับ .76 แสดงให้เห็นว่าความแปรปรวนที่อธิบายได้ด้วยโมเดลทั้งสองนี้ค่อนข้างสูง นั่นคือค่าความสามารถนักเรียนที่ประมาณได้จากโมเดลทั้งสองนี้ค่อนข้างคงเส้นคงวา และเมื่อพิจารณาค่าสถิติทดสอบความสอดคล้องกลมกลืนของข้อสอบรายข้อแบบถ่วงน้ำหนักพบว่าโมเดล UniFit-52D มีค่า MNSQ ระหว่าง .90 ถึง 1.10 (ดูตารางที่ B3.1D ในภาคผนวก) และโมเดล UniFit-52P มีค่า MNSQ ระหว่าง .90 ถึง 1.14 (ดูตารางที่ B3.1P ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์ขนาดอิทธิพลที่ยอมรับ ดังนั้นจึงไม่มีข้อสอบข้อใดแสดงความสอดคล้องกลมกลืนที่คลาดเคลื่อนจากข้อสอบข้ออื่นที่อยู่ในโมเดลเดียวกันและเป็นหลักฐานบ่งชี้ว่ามุมมองการวัดความสามารถทางวิทยาศาสตร์แบบภาพรวมสามารถนำมาใช้สร้างข้อสอบ ONET ปีการศึกษา 2552 ได้

ผลวิเคราะห์จากโมเดลพหุมิติระหว่างข้อสอบที่ให้คะแนนคำตอบแบบสองค่าหรือโมเดล BimFit-52D ในตารางที่ 4.7 แสดงให้เห็นว่าช่วงค่าความแปรปรวนกว้างระหว่าง .51 ถึง 1.24 นั่นคือความสามารถนักเรียนในแต่ละสาระการเรียนรู้ค่อนข้างแตกต่างกัน และเมื่อพิจารณาค่าความเที่ยงซึ่งอยู่ระหว่าง .48 ถึง .69 แสดงให้เห็นว่าโมเดล BimFit-52D สามารถอธิบายความแปรปรวนของความสามารถนักเรียนได้ในระดับปานกลาง ความสัมพันธ์ของค่าประมาณความสามารถนักเรียนของแต่ละสาระการเรียนรู้ในตารางที่ 4.8 ซึ่งมีค่าระหว่าง .47 ถึง .79 แสดงว่าในขณะที่ความสามารถนักเรียนแต่ละสาระการเรียนรู้สัมพันธ์กันในระดับปานกลางก็ยังคงมีความแตกต่างเกิดขึ้นและบ่งชี้ว่าเนื้อหาของข้อสอบของแต่ละสาระการเรียนรู้ไม่เหมือนกัน นอกจากนี้เมื่อพิจารณาความสอดคล้องกลมกลืนของข้อสอบแบบ infit พบว่ามีค่า MNSQ ระหว่าง .92 ถึง 1.13 (ดูตารางที่ B3.2D ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์สำคัญทางปฏิบัติจึงสรุปได้ว่าไม่มีหลักฐานบ่งชี้ว่าข้อสอบสอดคล้องกลมกลืนคลาดเคลื่อน นอกจากนี้ค่า MNSQ ที่อยู่ในช่วงเกณฑ์ดังกล่าวยังตีความได้ว่าไค์คุณลักษณะข้อสอบเชิงประจักษ์ที่สร้างจากคำตอบนักเรียนสามารถตรวจจับลักษณะไค์คุณลักษณะข้อสอบที่

สร้างขึ้นจากค่าประมาณความสามารถนักเรียนได้ดี นั่นหมายความว่าข้อสอบ ONET ปีการศึกษา 2552 สามารถจำแนกความสามารถที่แตกต่างกันของนักเรียนได้เช่นเดียวกับโมเดล UniFit-52D

ผลวิเคราะห์จากโมเดลพหุมิติภายในข้อสอบที่ให้คะแนนคำตอบแบบสองค่าทั้งประเภทที่กำหนดให้ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-allSt8-52D) และที่กำหนดให้ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-someSt8-52D) ในตารางที่ 4.7 พบว่า เมื่อกำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 ในโมเดล WimFit-allSt8-52D ค่าความแปรปรวนความสามารถนักเรียนของแต่ละสาระการเรียนรู้ต่ำกว่าโมเดล BimFit-52D อย่างชัดเจนและเมื่อกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ตามความเห็นของผู้เชี่ยวชาญในโมเดล WimFit-someSt8-52D พบว่าค่าความแปรปรวนค่อนข้างแตกต่างจากโมเดล BimFit-52D นอกจากนี้ความเที่ยงของสาระที่ 1 ถึง 7 ของโมเดล WimFit-allSt8-52D มีค่าน้อยมากเมื่อเปรียบเทียบกับโมเดล BimFit-52D และโมเดล WimFit-someSt8-52D แสดงให้เห็นว่าโมเดล WimFit-allSt8-52D อธิบายความแปรปรวนของความสามารถนักเรียนได้ต่ำและบ่งชี้ปัญหาการกำหนดลักษณะโมเดลคลาดเคลื่อนหรือข้อสอบทุกข้อไม่ได้วัดสาระที่ 8 ในขณะที่โมเดล WimFit-someSt8-52D ซึ่งผู้เชี่ยวชาญได้ประเมินว่ามีข้อสอบจำนวน 8 ข้อวัดสาระที่ 8 (ดูตารางที่ 3.5 ในบทที่ 3) ก็ไม่สามารถอธิบายความแปรปรวนของความสามารถนักเรียนในสาระนี้เพิ่มขึ้น แม้ว่าโมเดล WimFit-someSt8-52D จะมีความเที่ยงของสาระที่ 1 ถึง 7 สูงกว่าโมเดล WimFit-allSt8-52D ก็ตาม

ความสัมพันธ์ของค่าประมาณความสามารถในแต่ละสาระการเรียนรู้จากโมเดล WimFit-allSt8-52D มีค่าต่ำและติดลบโดยเฉพาะอย่างยิ่งกับสาระที่ 8 ซึ่งแตกต่างอย่างชัดเจนกับโมเดล WimFit-someSt8-52D ที่มีค่าบวกในระดับปานกลางยกเว้นสาระที่ 8 ที่ค่อนข้างต่ำ เมื่อพิจารณาค่าสหสัมพันธ์ร่วมกับค่าความแปรปรวนและความเที่ยงจึงพบว่าโมเดล WimFit-allSt8-52D ไม่เหมาะสมอย่างยิ่งสำหรับใช้ประมาณค่าความสามารถนักเรียน แม้ว่าค่า MNSQ ซึ่งอยู่ระหว่าง .92 ถึง 1.14 จะอยู่ในช่วงเกณฑ์ที่กำหนดก็ตาม นอกจากนี้แม้ว่าโมเดล WimFit-someSt8-52D จะมีค่าความเที่ยงสูงกว่าโมเดล WimFit-allSt8-52D แต่ก็มีค่าสหสัมพันธ์ของสาระที่ 8 ค่อนข้างต่ำด้วย ดังนั้นจึงไม่อาจกล่าวได้ว่าโมเดล WimFit-someSt8-52D เป็นโมเดลการวัดที่เหมาะสมที่สุดเช่นกัน

เมื่อเปลี่ยนวิธีให้คะแนนคำตอบเป็นแบบหลายค่า โมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-52P ในตารางที่ 4.7 มีค่าความแปรปรวนระหว่าง .39 ถึง .80 และมีค่าความเที่ยงอยู่ระหว่าง .46 ถึง .67 แสดงให้เห็นว่าโมเดล BimFit-52P สามารถอธิบายความแปรปรวนของความสามารถนักเรียนได้ในระดับปานกลาง ความสัมพันธ์ของค่าความสามารถนักเรียนของแต่ละสาระการเรียนรู้ในตารางที่ 4.8 ซึ่งมีค่าระหว่าง .44 ถึง .79 แสดงให้เห็นว่านักเรียนมีความสามารถแตกต่างกันในแต่ละสาระการเรียนรู้ นอกจากนี้เมื่อพิจารณาความสอดคล้องกลมกลืนของข้อสอบแบบ infit พบว่ามีค่า MNSQ ระหว่าง .92 ถึง 1.14 (ดูตารางที่ B3.2P ในภาคผนวก) ซึ่งอยู่ในช่วงเกณฑ์คุณภาพที่กำหนด จึงสรุปได้ว่าไม่มีหลักฐานบ่งชี้ว่าข้อสอบสอดคล้องกลมกลืนคลาดเคลื่อน นอกจากนี้ค่า MNSQ ที่อยู่ในช่วงเกณฑ์ดังกล่าวยังสามารถตีความได้ว่าคงคุณลักษณะข้อสอบเชิงประจักษ์ที่สร้างจากคำตอบ

นักเรียนสามารถตรวจจับลักษณะโค้งคุณลักษณะข้อสอบที่สร้างขึ้นจากค่าประมาณความสามารถนักเรียนได้ดี นั่นหมายความว่าเมื่อให้คะแนนคำตอบข้อสอบ ONET ปีการศึกษา 2552 แบบหลายค่าข้อสอบสามารถจำแนกความสามารถที่แตกต่างกันของนักเรียนได้

ผลวิเคราะห์จากโมเดลพหุมิติภายในข้อสอบที่ให้คะแนนคำตอบแบบหลายค่าทั้งประเภทที่กำหนดให้ข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-allSt8-52P) และที่กำหนดให้ข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8 (โมเดล WimFit-someSt8-52P) ในตารางที่ 4.7 พบว่า เมื่อกำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 ในโมเดล WimFit-allSt8-52P ค่าความแปรปรวนความสามารถนักเรียนของแต่ละสาระการเรียนรู้ต่ำกว่าโมเดล BimFit-52P อย่างชัดเจนและเมื่อกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ตามความเห็นของผู้เชี่ยวชาญโมเดล WimFit-someSt8-52P พบว่าค่า

ตารางที่ 4.7 ค่าความแปรปรวนและความเที่ยงของผลการวัดด้วยโมเดลพหุมิติ ปีการศึกษา 2552

สาระการเรียนรู้	โมเดลพหุมิติระหว่างข้อสอบ		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8		โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8	
	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV	ความแปรปรวน	ความเที่ยง EAP/PV
ให้คะแนนคำตอบแบบสองค่า						
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.61	.68	.28	.12	.57	.66
2: ชีวิตกับสิ่งแวดล้อม	1.24	.51	.68	.18	.88	.53
3: สารและสมบัติของสาร	.61	.69	.30	.12	.58	.62
4: แรงและการเคลื่อนที่	.82	.63	.37	.16	.75	.64
5: พลังงาน	.77	.62	.38	.15	.75	.63
6: กระบวนการเปลี่ยนแปลงของโลก	.75	.49	.49	.13	.72	.53
7: ดาราศาสตร์และอวกาศ	.52	.48	.40	.14	.43	.49
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-	-	.53	.57	.20	.21
ให้คะแนนคำตอบแบบหลายค่า						
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.50	.66	.26	.09	.43	.65
2: ชีวิตกับสิ่งแวดล้อม	.77	.48	.53	.15	.57	.49
3: สารและสมบัติของสาร	.61	.67	.32	.15	.51	.63
4: แรงและการเคลื่อนที่	.53	.59	.28	.09	.44	.61
5: พลังงาน	.48	.59	.32	.10	.46	.60
6: กระบวนการเปลี่ยนแปลงของโลก	.80	.48	.49	.14	.64	.52
7: ดาราศาสตร์และอวกาศ	.39	.46	.37	.15	.30	.50
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-	-	.46	.53	.19	.27

ตารางที่ 4.8 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2552

สาระการเรียนรู้	สัมประสิทธิ์สหสัมพันธ์						
	1	2	3	4	5	6	7
โมเดลพหุมิติระหว่างข้อสอบ (ให้คะแนนสองค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.66						
3: สารและสมบัติของสาร	.79	.64					
4: แรงและการเคลื่อนที่	.74	.59	.76				
5: พลังงาน	.74	.59	.74	.72			
6: กระบวนการเปลี่ยนแปลงของโลก	.64	.52	.64	.56	.58		
7: ดาราศาสตร์และอวกาศ	.56	.52	.60	.56	.60	.47	
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบทุกข้อวัดสาระที่ 8 (ให้คะแนนสองค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.22						
3: สารและสมบัติของสาร	.52	.18					
4: แรงและการเคลื่อนที่	.39	.14	.44				
5: พลังงาน	.38	.16	.37	.35			
6: กระบวนการเปลี่ยนแปลงของโลก	.27	.09	.23	.18	.18		
7: ดาราศาสตร์และอวกาศ	.27	.07	.30	.20	.21	.16	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-.27	.10	-.23	-.06	-.07	-.19	-.44
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบบางข้อวัดสาระที่ 8 (ให้คะแนนสองค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.71						
3: สารและสมบัติของสาร	.79	.65					
4: แรงและการเคลื่อนที่	.77	.64	.76				
5: พลังงาน	.74	.62	.70	.73			
6: กระบวนการเปลี่ยนแปลงของโลก	.67	.57	.64	.64	.65		
7: ดาราศาสตร์และอวกาศ	.63	.56	.61	.61	.61	.51	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	.13	.18	.04	.20	.28	.19	.22
โมเดลพหุมิติระหว่างข้อสอบ (ให้คะแนนหลายค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.60						
3: สารและสมบัติของสาร	.79	.58					
4: แรงและการเคลื่อนที่	.71	.54	.73				
5: พลังงาน	.71	.55	.71	.68			
6: กระบวนการเปลี่ยนแปลงของโลก	.61	.48	.62	.54	.54		
7: ดาราศาสตร์และอวกาศ	.58	.48	.56	.53	.54	.44	

ตารางที่ 4.8 ค่าสัมประสิทธิ์สหสัมพันธ์ความสามารถนักเรียน ปีการศึกษา 2552 (ต่อ)

สาระการเรียนรู้	สัมประสิทธิ์สหสัมพันธ์						
	1	2	3	4	5	6	7
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบทุกข้อวัดสาระที่ 8 (ให้คะแนนหลายค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.31						
3: สารและสมบัติของสาร	.52	.27					
4: แรงและการเคลื่อนที่	.46	.19	.43				
5: พลังงาน	.42	.26	.37	.39			
6: กระบวนการเปลี่ยนแปลงของโลก	.30	.16	.31	.22	.24		
7: ดาราศาสตร์และอวกาศ	.42	.25	.29	.32	.32	.17	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-.33	-.18	-.16	-.28	-.29	-.14	-.55
โมเดลพหุมิติภายในข้อสอบที่ข้อสอบบางข้อวัดสาระที่ 8 (ให้คะแนนหลายค่า)							
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต							
2: ชีวิตกับสิ่งแวดล้อม	.66						
3: สารและสมบัติของสาร	.79	.64					
4: แรงและการเคลื่อนที่	.76	.58	.77				
5: พลังงาน	.73	.60	.72	.72			
6: กระบวนการเปลี่ยนแปลงของโลก	.68	.55	.68	.64	.64		
7: ดาราศาสตร์และอวกาศ	.68	.57	.66	.62	.61	.54	
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	.26	.26	.21	.30	.36	.29	.23

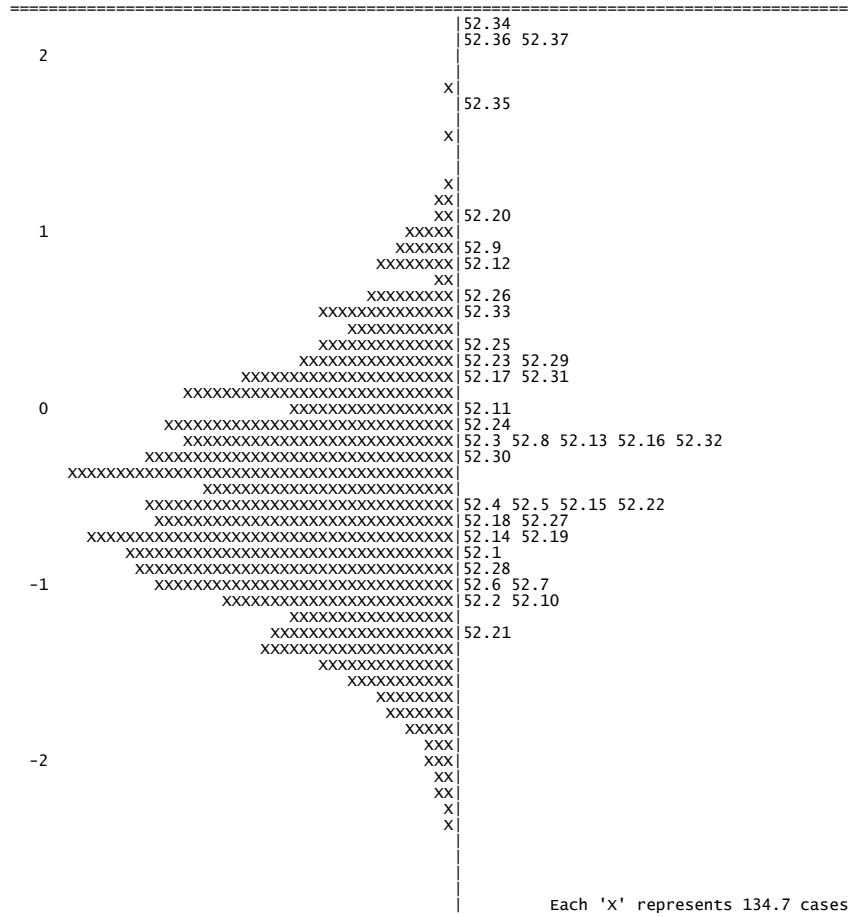
ความแปรปรวนค่อนข้างแตกต่างจากโมเดล BimFit-52P นอกจากนี้ความเที่ยงของสาระที่ 1 ถึง 7 ของโมเดล WimFit-allSt8-52P มีค่าน้อยมากเมื่อเปรียบเทียบกับโมเดล BimFit-52P และโมเดล WimFit-someSt8-52P แสดงให้เห็นว่าโมเดล WimFit-allSt8-52P อธิบายความแปรปรวนของความสามารถนักเรียนได้ต่ำและบ่งชี้ปัญหาข้อกำหนดลักษณะโมเดลคลาดเคลื่อนหรือข้อสอบทุกข้อไม่ได้วัดสาระที่ 8 ในขณะที่โมเดล WimFit-someSt8-52P ซึ่งผู้เชี่ยวชาญได้ประเมินว่ามีข้อสอบจำนวน 8 ข้อวัดสาระที่ 8 (ดูตารางที่ 3.5 ในบทที่ 3) มีค่าความเที่ยงในภาพรวมสูงกว่าโมเดล BimFit-52P แสดงให้เห็นว่าโมเดล WimFit-someSt8-52P อธิบายความแปรปรวนของความสามารถนักเรียนได้ดีกว่าโมเดล BimFit-52P และโมเดล WimFit-allSt8-52P

ความสัมพันธ์ของค่าความสามารถของแต่ละสาระการเรียนรู้จากโมเดล WimFit-allSt8-52P มีค่าลบในสาระที่ 8 ซึ่งแตกต่างอย่างชัดเจนกับโมเดล WimFit-someSt8-52P ที่มีค่าบวก เมื่อพิจารณาค่าสหสัมพันธ์ร่วมกับค่าความแปรปรวนและความเที่ยงจึงพบว่าโมเดล WimFit-allSt8-52P ไม่เหมาะสมอย่างยิ่งต่อการประมาณค่าความสามารถนักเรียน แม้ว่าค่า MNSQ ซึ่งอยู่ระหว่าง .92 ถึง 1.13 จะอยู่ในช่วงเกณฑ์ที่กำหนดก็ตาม

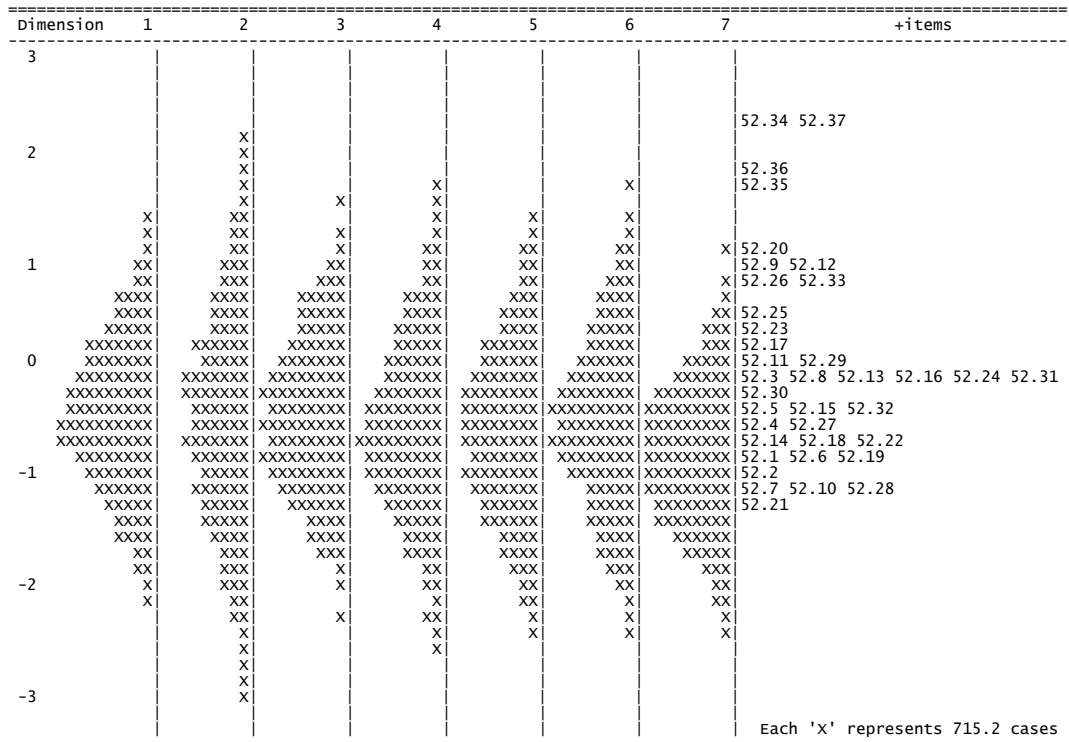
แผนที่ของ Wright ในภาพที่ 4.3(a) ได้จากการประมาณด้วยโมเดลเอกมิติ UniFit-52D เมื่อพิจารณาค่าความยากข้อสอบพบว่ามีค่าระหว่าง -1.22 ถึง 2.22 ลอจิทซึ่งมีพิสัยใกล้เคียงกับกับค่าความสามารถนักเรียนที่มีค่าระหว่าง -2.01 ถึง 1.67 ลอจิทและมีข้อที่ 34 ยากที่สุดและข้อที่ 21 ง่ายที่สุด เมื่อพิจารณาการแจกแจงค่าความยากข้อสอบที่อยู่สูงกว่าการแจกแจงค่าความสามารถนักเรียนอย่างเห็นได้ชัด แสดงว่าโดยภาพรวมข้อสอบ ONET เป็นข้อสอบยาก ตำแหน่งความยากข้อสอบกระจายมากกว่า ข้อสอบส่วนใหญ่จึงบรรยายนักเรียนที่มีความสามารถสูงได้ดี อย่างไรก็ตามไม่มีข้อสอบข้อใดเลยที่เหมาะสมกับนักเรียนที่มีค่าความสามารถต่ำกว่า -1.22 ลอจิท นั่นคือข้อสอบ ONET ยังไม่สามารถบรรยายความสามารถนักเรียนกลุ่มนี้ได้

เมื่อพิจารณาแผนที่ของ Wright สำหรับโมเดลพหุมิติระหว่างข้อสอบ BimFit-52D ในภาพที่ 4.3(b) พบว่า การแจกแจงค่าความยากข้อสอบอยู่สูงกว่าการแจกแจงความสามารถนักเรียนในทุกระยะการเรียนรู้ซึ่งแสดงให้เห็นว่านักเรียนส่วนใหญ่มีความสามารถในแต่ละระยะการเรียนรู้ใกล้เคียงกันและอยู่ในระดับค่อนข้างต่ำถึงปานกลางเมื่อเปรียบเทียบกับความยากข้อสอบ นอกจากนี้เมื่อพิจารณาการแจกแจงค่าความสามารถนักเรียนในสาระที่ 2 ซึ่งมีลักษณะแบนแสดงว่านักเรียนมีความสามารถหลากหลายมากกว่าสาระการเรียนรู้ที่เหลือซึ่งมีลักษณะการแจกแจงโด่งกว่า เมื่อพิจารณาค่าความยากข้อสอบของโมเดล BimFit-52D ซึ่งอยู่ระหว่าง -1.33 ถึง 2.33 ลอจิทพบว่ามีช่วงกว้างกว่าโมเดล UniFit-52D และมีข้อที่ 34 เป็นข้อที่ยากที่สุด ในขณะที่ข้อที่ 21 เป็นข้อที่ง่ายที่สุด

แผนที่ของ Wright ในภาพที่ 4.3(c) ได้จากโมเดลที่กำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 หรือโมเดล WimFit-allSt8-52 ซึ่งค่าความยากข้อสอบอยู่ในช่วงแคบระหว่าง -.39 ถึง 1.44 เมื่อเปรียบเทียบกับลักษณะการแจกแจงค่าความสามารถนักเรียน โดยมีข้อที่ 21 เป็นข้อที่ง่ายที่สุดและมีข้อที่ 34 เป็นข้อที่ยากที่สุดและแสดงให้เห็นว่าข้อสอบส่วนใหญ่ค่อนข้างยากเมื่อเปรียบเทียบกับลักษณะการแจกแจงค่าความสามารถ ส่วนภาพที่ 4.3(d) เป็นแผนที่ของ Wright ที่ได้จากโมเดล WimFit-someSt8-52D ซึ่งกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ค่าความยากข้อสอบที่ได้จากโมเดลนี้อยู่ระหว่าง -.77 ถึง 2.85 และมีข้อที่ 34 เป็นข้อที่ยากที่สุดและข้อ 21 เป็นข้อที่ง่ายที่สุด ข้อสังเกตระหว่างแผนที่ Wright จากโมเดล WimFit-allSt8-52 และ WimFit-someSt8-52 คือเมื่อกำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 ตามความเห็นของผู้เชี่ยวชาญ การแจกแจงค่าความสามารถนักเรียนในสาระที่ 1 ถึง 7 มีลักษณะแบน ในขณะที่สาระที่ 8 มีการแจกแจงแบบโด่ง นั้นหมายความว่าข้อสอบวัดสาระที่ 8 สามารถจำแนกความสามารถได้ดี แต่กลับทำให้การจำแนกความสามารถนักเรียนในสาระอื่นลดลงรวมทั้งจะเห็นได้ว่าช่วงค่าความยากข้อสอบกว้างเพียงพอที่จะครอบคลุมการแจกแจงความสามารถนักเรียนส่วนใหญ่ นั้นหมายความว่าแม้ว่าผลการประมาณค่าความยากข้อสอบด้วยโมเดล WimFit-someSt8-52 จะทำให้ข้อสอบยากขึ้น แต่ช่วงความกว้างของก็สามารถบรรยายความสามารถของนักเรียนได้ดีกว่าโมเดล WimFit-allSt8-52 ที่ให้ค่าความยากข้อสอบเกาะกลุ่มกันและไม่สามารถบรรยายความสามารถนักเรียนบางกลุ่มได้



(a) โมเดลเอกมิตินี้ให้คะแนนคำตอบสองค่า



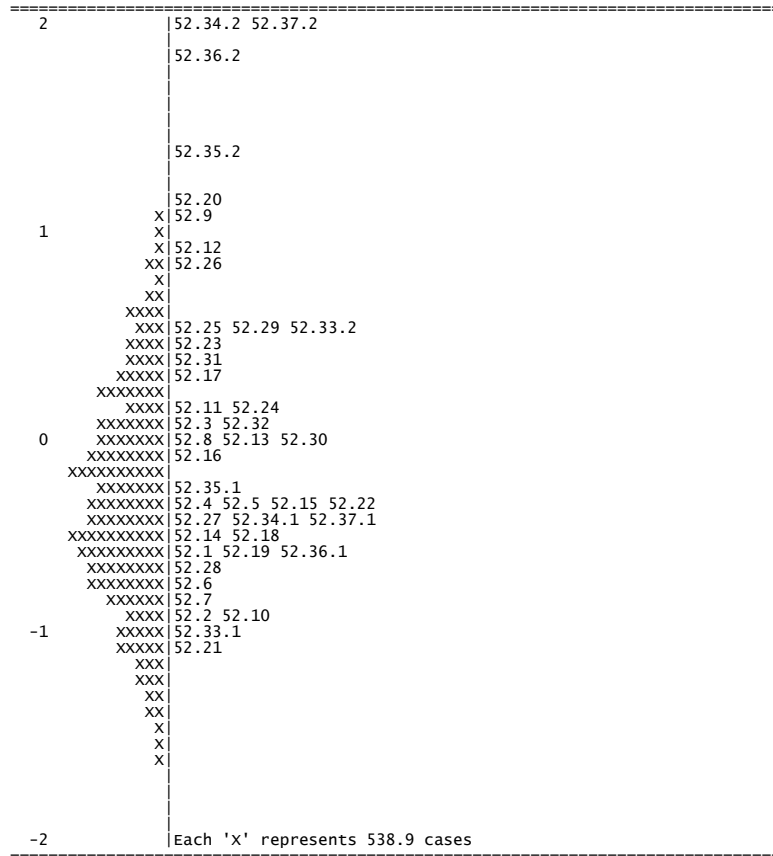
(b) โมเดลพหุมิตีระหว่างข้อสอบที่ให้คะแนนคำตอบสองค่า

[illegible]

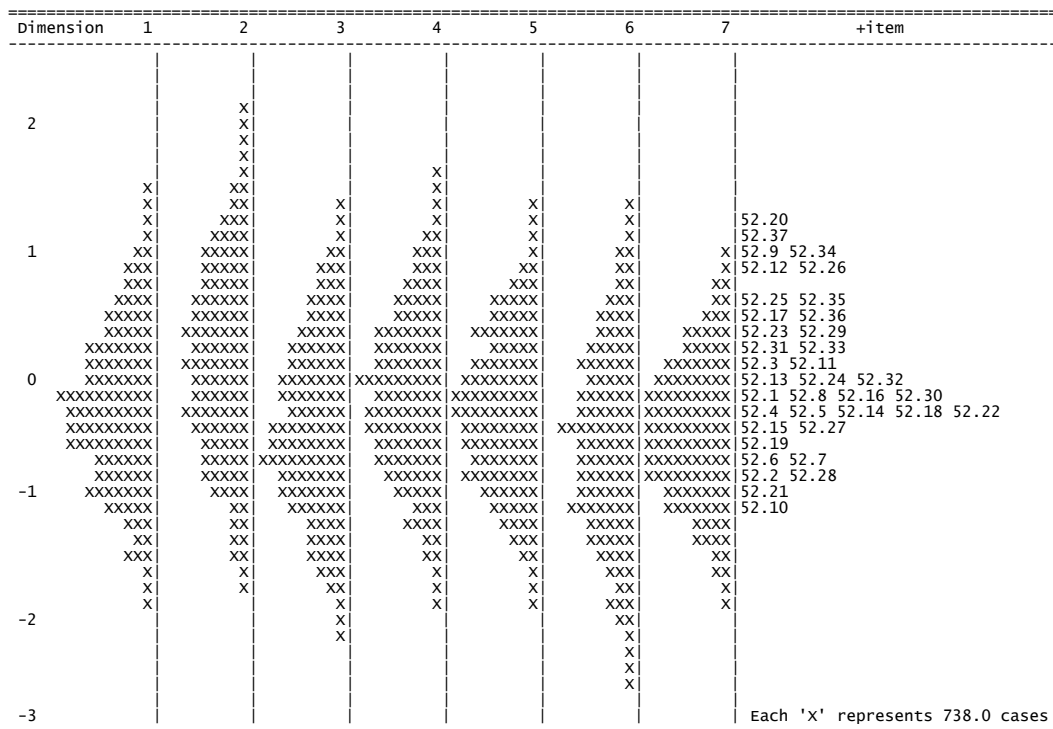
(c) โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนคำตอบสองค่าโดยข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8

Dimension	1	2	3	4	5	6	7	8	+items
3									52.34
									52.36 52.37
									52.35
2		X		X					
		X		X	X	X			
	X	X	X	X	X	X	X		52.20
	XX	XX	X	X	XX	XX			52.9
	XX	XX	XX	XX	XX	XX	X		52.12
	XX	XX	XX	XX	XX	XX	X		52.26 52.33
1	XX	XX	XX	XX	XX	XX	XX	X	
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.23 52.25 52.29
	XXX	XXX	XXXX	XXXX	XXX	XXXX	XXX	XX	52.17 52.31
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.24
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.8 52.16 52.30 52.32
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.3 52.11 52.13
0	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.4 52.5 52.15 52.22 52.27
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.14 52.18
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.1 52.7 52.10 52.19
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.6 52.28
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.2
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.21
-1	XXX	XX	XX	XX	XX	XX	XX	X	
	XX	XX	XX	XX	XX	XX	X		
	X	XX	X	X	X	XX	X		
	X	XX	X	X	X	X	X		
		X		X	X	X			
-2		X		X		X			Each 'x' represents BIG cases

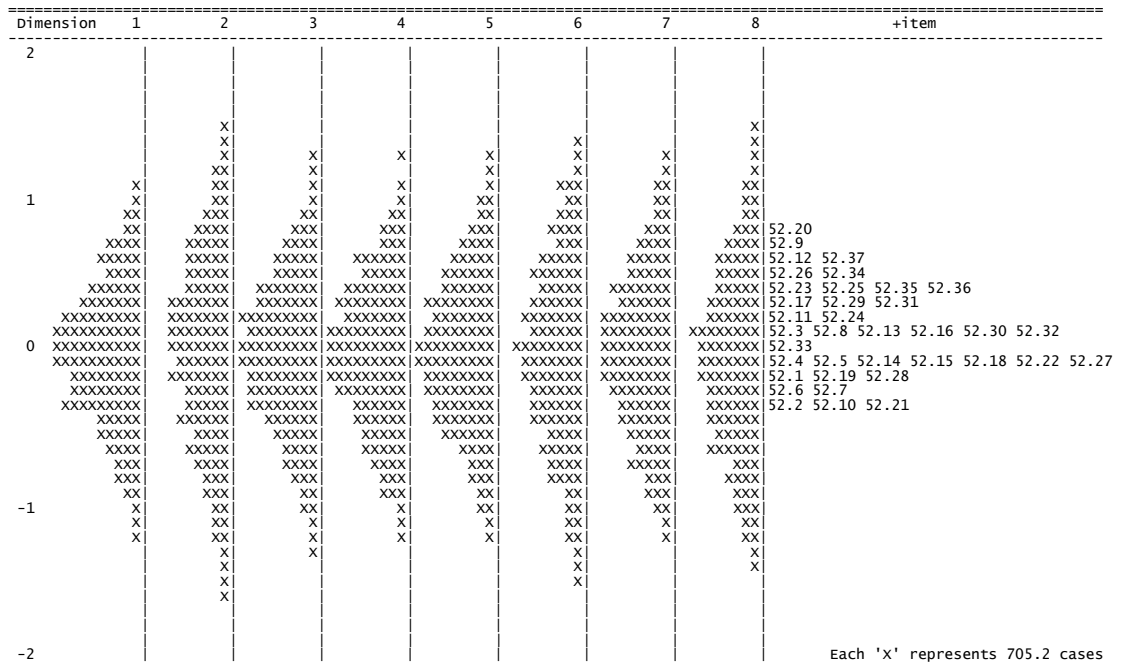
(d) โมเดลพหุนิติภายในข้อสอบที่ให้คะแนนคำตอบสองค่าโดยข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8



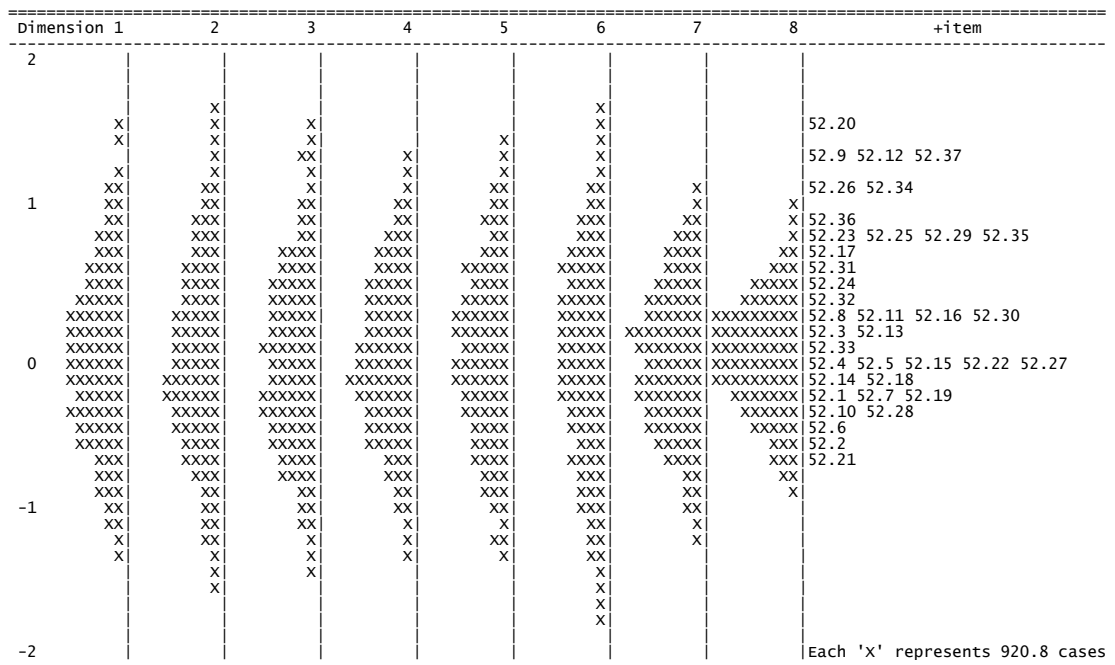
(e) โมเดลเอกมิติตีที่ให้คะแนนคำตอบแบบหลายค่า



(f) โมเดลพหุมิตีระหว่างข้อสอบที่ให้คะแนนคำตอบแบบหลายค่า



(g) โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนคำตอบแบบหลายค่าโดยข้อสอบทุกข้อวัดสาระการเรียนรู้ที่ 8



(h) โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนคำตอบแบบหลายค่าโดยข้อสอบบางข้อวัดสาระการเรียนรู้ที่ 8

ภาพที่ 4.3 แผนที่ของ Wright จากโมเดลการวัด ปีการศึกษา 2552

เมื่อเปลี่ยนวิธีให้คะแนนคำตอบเป็นแบบหลายค่า การแจกแจงค่าความยากข้อสอบในแผนที่ของ Wright ในภาพที่ 4.3(e) แสดงให้เห็นว่าโมเดลเอกมิติที่ให้คะแนนคำตอบมากกว่าสองค่า UniFit-52P ให้สารสนเทศในภาพรวมค่อนข้างใกล้เคียงกับโมเดล UniFit-52D แต่อย่างไรก็ตาม ข้อสอบมีตำแหน่งค่าความยากแยกออกจากกันชัดเจนมากขึ้น โดยข้อสอบที่เป็นข้อยากจะมีตำแหน่งสูงกว่าระดับปานกลางอย่างเห็นได้ชัด โดยเฉพาะข้อสอบที่กำหนดให้นักเรียนคำตอบสองตัวเลือกจากหกตัวเลือก (ข้อที่ 34 ถึง 37) นั้นหมายความว่านักเรียนที่เลือกคำตอบได้ถูกต้องครบถ้วนทั้งสองตัวเลือกในแต่ละข้อเป็นนักเรียนที่มีความสามารถสูงกว่านักเรียนที่ตอบถูกเพียงตัวเลือกเดียว นอกจากนี้การแจกแจงค่าความยากข้อสอบก็มีขอบเขตกว้างและครอบคลุมช่วงการแจกแจงค่าความสามารถนักเรียนเกือบทั้งหมด แสดงว่าการให้คะแนนคำตอบมากกว่าสองค่าจะช่วยทำให้ข้อสอบสามารถบรรยายความสามารถของนักเรียนได้ดียิ่งขึ้น

ตารางที่ 4.9 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดล ปีการศึกษา 2552

โมเดล	จำนวนพารามิเตอร์	AIC	BIC	G ²
ให้คะแนนคำตอบแบบสองค่า				
โมเดลเอกมิติ	38	3857835.92	3857947.81	3857759.92
โมเดลพหุมิติระหว่างข้อสอบ	65	3862916.54	3863107.93	3862786.54
โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8	73	3862052.44	3862267.39	3861906.44
โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8	73	3859428.41	3859643.36	3859282.41
ให้คะแนนคำตอบแบบหลายค่า				
โมเดลเอกมิติ	43	4391615.05	4391741.66	4391529.05
โมเดลพหุมิติระหว่างข้อสอบ	70	4397208.83	4397414.94	4397068.83
โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8	78	4397460.13	4397689.80	4397304.13
โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8	78	4389833.18	4390062.85	4389677.18

เมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.3(f) ซึ่งได้จากโมเดลพหุมิติระหว่างข้อสอบที่ให้คะแนนคำตอบแบบหลายค่าหรือโมเดล BimFit-52P พบว่า ตำแหน่งค่าความยากข้อสอบอยู่ใกล้กันและเป็นข้อสอบที่เหมาะสมกับนักเรียนที่มีความสามารถปานกลางไปจนถึงนักเรียนที่มีความสามารถสูง ข้อสอบสามารถบรรยายความสามารถของนักเรียนได้ละเอียดทุกระดับความสามารถ แต่ก็ส่งผลให้ข้อสอบที่วัดสาระที่ 2 และ 6 จำแนกความสามารถนักเรียนลดลงเมื่อเปรียบเทียบกับการให้คะแนนคำตอบแบบสองค่า เมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.3(g) และ 4.3(h) ซึ่งได้จากโมเดล พหุมิติภายในข้อสอบที่กำหนดให้ข้อสอบทุกข้อวัดสาระที่ 8 หรือโมเดล WimFit-allSt8-52P และข้อสอบ

บางข้อวัดสาระที่ 8 หรือโมเดล WimFit-someSt8-52P ตามลำดับ พบว่าช่วงการแจกแจงค่าความยากข้อสอบของโมเดล WimFit-allSt8-52P แคบและมีตำแหน่งค่าความยากข้อสอบครอบคลุมความสามารถนักเรียนน้อยกว่าลักษณะการแจกแจงค่าความยากข้อสอบของโมเดล WimFit-someSt8-52P นั้นหมายความว่าโมเดล WimFit-someSt8-52P สามารถบรรยายความสามารถนักเรียนที่แตกต่างกันได้ดีกว่า

ตารางที่ 4.9 แสดงโมเดลเอกมิตีและโมเดลพหุมิตีจำนวนสองชุด โดยแต่ละชุดมีวิธีให้คะแนนคำตอบแตกต่างกันตามที่ผู้วิจัยได้อธิบายผลวิเคราะห์มาตั้งแต่ต้นของเนื้อหาในส่วนนี้ ข้อสังเกตที่สำคัญเพื่อป้องกันความเข้าใจคลาดเคลื่อนเมื่อข้อมูลชุดเดียวกันแต่มีวิธีให้คะแนนคำตอบต่างกันคือ ข้อมูลคะแนนคำตอบที่ให้คะแนนแบบสองค่าเป็นข้อมูลคนละชุดกับข้อมูลที่ให้คะแนนคำตอบแบบหลายค่า แม้ว่ากลุ่มตัวอย่างจะเป็นกลุ่มเดียวกันและทำข้อสอบ ONET เหมือนกันก็ตาม ดังนั้นหลักการเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลการวัดต่างๆ จึงไม่สามารถนำโมเดลที่ได้จากข้อมูลคนละชุดกันมาเปรียบเทียบกันได้

เมื่อพิจารณาโมเดลที่ให้คะแนนคำตอบแบบสองค่าพบว่า โมเดลเอกมิตี UniFit-52D มีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากับ 38 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 36 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละหนึ่งค่า สำหรับโมเดลพหุมิตีระหว่างข้อสอบ BimFit-52D จะมีพารามิเตอร์ที่ต้องประมาณจำนวน 65 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 30 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 7 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 21 ค่า และสำหรับโมเดลพหุมิตีภายในข้อสอบทั้งโมเดล WimFit-allSt8-52D และโมเดล WimFit-someSt8-52D จะมีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากันคือ 73 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 29 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 8 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 28 ค่า เมื่อเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลที่ให้คะแนนคำตอบแบบสองค่า พบว่าโมเดลเอกมิตีสอดคล้องกลมกลืนกับข้อมูลดีกว่าโมเดลพหุมิตีทั้งหมดเนื่องจากมีค่า G^2 น้อยที่สุดและผลต่างของค่า G^2 มีนัยสำคัญทางสถิติ (โมเดล Unifit-52D กับ BimFit-52D: $\chi^2=5,026.62$, $df=27$, $p<.01$; โมเดล UniFit-52D กับ WimFit-allSt8-52D: $\chi^2=4,146.52$, $df=35$, $p<.01$; โมเดล UniFit-52D กับ WimFit-someSt8-52D: $\chi^2=1,522.49$, $df=35$, $p<.01$) และเมื่อพิจารณาค่า AIC และ BIC ของโมเดลเอกมิตีพบว่ามีความน้อยที่สุดเช่นกัน แสดงให้เห็นว่าโมเดลเอกมิตีเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด

เมื่อพิจารณาโมเดลที่ให้คะแนนคำตอบแบบหลายค่าพบว่า โมเดลเอกมิตี UniFit-52P มีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากับ 43 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 41 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละหนึ่งค่า สำหรับโมเดลพหุมิตีระหว่างข้อสอบ BimFit-52P จะมีพารามิเตอร์ที่ต้องประมาณจำนวน 70 ค่า

ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 35 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 7 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 21 ค่า และสำหรับโมเดลพหุมิติภายในข้อสอบทั้งโมเดล WimFit-allSt8-52P และโมเดล WimFit-someSt8-52P จะมีจำนวนพารามิเตอร์ที่ต้องประมาณเท่ากันคือ 78 ค่า ประกอบด้วยพารามิเตอร์ความยากข้อสอบจำนวน 34 ค่า ค่าเฉลี่ยและค่าความแปรปรวนความสามารถของประชากรนักเรียนจำนวนอย่างละ 8 ค่า และความแปรปรวนรวมความสามารถของประชากรนักเรียนระหว่างสาระการเรียนรู้จำนวน 28 ค่า เมื่อเปรียบเทียบความสอดคล้องกลมกลืนของโมเดลที่ให้คะแนนคำตอบแบบสองค่า พบว่าโมเดล WimFit-someSt8-52P สอดคล้องกลมกลืนกับข้อมูลดีกว่าโมเดลการอื่นๆ เนื่องจากมีค่า G^2 น้อยที่สุดและผลต่างของค่า G^2 มีนัยสำคัญทางสถิติ (โมเดล WimFit-someSt8-52P กับ UniFit-52P: $\chi^2=1,851.87$, $df=27$, $p<.01$; โมเดล WimFit-someSt8-52P กับ BimFit-52P: $\chi^2=7,391.65$, $df=35$, $p<.01$; โมเดล WimFit-someSt8-52P กับ WimFit-allSt8-52P: $\chi^2=7,626.95$, $df=35$, $p<.01$) และเมื่อพิจารณาค่า AIC และ BIC ของโมเดลเอกมิติพบว่ามีค่าน้อยที่สุดเช่นกัน แสดงให้เห็นว่าโมเดลพหุมิติภายในข้อสอบที่กำหนดให้ข้อสอบบางข้อวัดสาระที่ 8 เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด

ตอนที่ 2 ผลการตรวจสอบความตรงข้ามกลุ่มของโมเดลการวัด

วัตถุประสงค์หลักของการตรวจสอบความตรงข้ามกลุ่ม (cross validation) ของโมเดลคือ เพื่อแสดงหลักฐานเชิงประจักษ์สนับสนุนคุณสมบัติความไม่แปรเปลี่ยน (invariance) ของโมเดลที่ประมาณค่าพารามิเตอร์เชิงโครงสร้างจากกลุ่มตัวอย่างหนึ่งได้ใกล้เคียงกับอีกกลุ่มตัวอย่างหนึ่งซึ่งโครงการวิจัยนี้นอกจากจะกำหนดให้กลุ่มตัวอย่างทั้งสองกลุ่มถูกสุ่มอย่างง่ายจากฐานข้อมูลวิเคราะห์เดียวกันแล้วยังกำหนดให้มีขนาดเท่ากันด้วย (ดูตารางที่ 3.1 หน้า 48)

ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลคำตอบข้อสอบ ONET ของแต่ละปีการศึกษาพบว่า โมเดลที่สอดคล้องกลมกลืนมากที่สุดของปีการศึกษา 2550 คือโมเดลพหุมิติระหว่างข้อสอบ ปีการศึกษา 2551 คือโมเดลเอกมิติ และปีการศึกษา 2552 คือโมเดลเอกมิติเมื่อให้คะแนนคำตอบแบบสองค่า และเป็นโมเดลพหุมิติภายในข้อสอบเมื่อให้คะแนนแบบหลายค่า ผลวิเคราะห์ดังกล่าวแสดงให้เห็นว่าข้อสอบ ONET แต่ละปีการศึกษาไม่เหมือนกันทั้งในด้านคุณสมบัติที่สังเกตได้ภายนอก เช่นสัดส่วนจำนวนข้อสอบที่วัดในแต่ละสาระการเรียนรู้ รูปแบบการตอบข้อสอบ และลักษณะโจทย์คำถาม รวมทั้งยังมีคุณสมบัติแฝงที่ไม่เหมือนกันด้วย ได้แก่ค่าพารามิเตอร์ความยากข้อสอบ ค่าเฉลี่ยความสามารถประชากรนักเรียนและความแปรปรวนความสามารถของประชากรนักเรียน ดังนั้นเพื่อให้กระบวนการตรวจสอบความตรงข้ามกลุ่มสอดคล้องกับกรอบการวิจัยในครั้งนี้ ผู้วิจัยจึงกำหนดหลักเกณฑ์พิจารณาเลือกโมเดลสำหรับตรวจสอบความตรงข้ามกลุ่มดังนี้

สำหรับปีการศึกษา 2550 ผู้วิจัยเลือกโมเดลพหุมิติระหว่างข้อสอบ BimFit-50 สำหรับตรวจสอบความตรงข้ามกลุ่มเป็นโมเดลแรก ข้อดีของโมเดลนี้นอกจากจะสอดคล้องกลมกลืนกับข้อมูลคำตอบมากกว่าโมเดลพหุมิติอื่นๆ แล้ว ยังรักษาความตั้งใจเดิมที่ผู้สร้างข้อสอบ ONET ต้องการให้ข้อสอบวัดความสามารถนักเรียนแยกระายสาระการเรียนรู้ นั่นคือผู้ใช้คะแนน ONET จะได้สารสนเทศความสามารถนักเรียนชัดเจนมากขึ้นและตรงกับความต้องการของผู้สร้างแบบสอบหรือเรียกว่ามีความตรงตามจุดมุ่งหมาย (consequential validity) แต่อย่างไรก็ตาม โมเดล BimFit-50 มีข้อจำกัดสองประการคือให้สารสนเทศความสามารถนักเรียนไม่ครบตามโครงสร้างสาระการเรียนรู้ในหลักสูตรแกนกลางฯ พ.ศ. 2551 โดยไม่มีค่าประมาณความสามารถนักเรียนในสาระที่ 8 เนื่องจากโมเดลพหุมิติภายในข้อสอบทั้งสองโมเดลสอดคล้องกลมกลืนกับข้อมูลน้อยกว่า รวมทั้งมีค่าสหสัมพันธ์ความสามารถนักเรียนรายสาระการเรียนรู้ต่ำและติดลบ (ดูตารางที่ 4.2) ยิ่งแสดงว่าโมเดลพหุมิติภายในข้อสอบทั้งสองโมเดลให้ผลวิเคราะห์ขัดแย้งกับหลักการให้คะแนนคำตอบ กล่าวคือเมื่อนักเรียนตอบข้อสอบที่วัดสาระที่ 8 ถูกจะหมายถึงตอบถูกในสาระอื่นที่ข้อสอบข้อนั้นวัดด้วย การได้คะแนนในแต่ละข้อจึงหมายถึงนักเรียนบรรลุทั้งสองสาระโดยอัตโนมัติ ดังนั้นค่าสหสัมพันธ์จึงไม่ควรติดลบ ข้อจำกัดประการที่สองคือโมเดล BimFit-50 ไม่สามารถให้สารสนเทศความสามารถนักเรียนในภาพรวมได้ตรงตามวัตถุประสงค์ของการวิจัย และการหาผลรวมจากค่าความสามารถในแต่ละสาระที่ได้จากโมเดล BimFit-50 ไม่สามารถทำได้เช่นกัน เนื่องจากกรอบโมเดล MRCML จะกำหนดให้ค่าเฉลี่ยพารามิเตอร์ความยากข้อสอบของแต่ละสาระการเรียนรู้มีค่าเท่ากับศูนย์ ดังนั้นโมเดลที่ผู้วิจัยเลือกใช้เป็นโมเดลที่สองสำหรับตรวจสอบความตรงข้ามกลุ่มของปีการศึกษา 2550 คือโมเดลเอกมิติ UniFit-50 แม้ว่าโมเดลดังกล่าวจะสอดคล้องกลมกลืนกับข้อมูลน้อยที่สุดก็ตาม

สำหรับปีการศึกษา 2551 ผู้วิจัยเลือกโมเดลเอกมิติ UniFit-51 เป็นโมเดลแรกสำหรับตรวจสอบความตรงข้ามกลุ่ม เนื่องจากสอดคล้องกลมกลืนกับข้อมูลคำตอบมากที่สุด ข้อดีของโมเดลนี้คือใช้คะแนนรวมจากข้อสอบทุกข้อประมาณค่าความสามารถนักเรียนทำให้ได้สารสนเทศในภาพรวมสำหรับประเมินผลสัมฤทธิ์ทางการเรียน (summative assessment) รวมทั้งเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดขั้นต้นในภาพรวมด้วย แต่อย่างไรก็ตามโมเดล UniFit-51 มีข้อจำกัดคือฝ่าฝืนธรรมชาติของความสามารถนักเรียนที่ใช้ทำข้อสอบและความตั้งใจของผู้สร้างข้อสอบ ONET ที่ต้องการใช้ข้อสอบ ONET เป็นเครื่องมือบ่งชี้ความสามารถนักเรียนรายสาระการเรียนรู้ และไม่สามารถให้สารสนเทศสำหรับกำหนดคะแนนจุดตัดขั้นต้นรายสาระการเรียนรู้ได้ ดังนั้นผู้วิจัยจึงพิจารณาเลือกโมเดลที่สองคือโมเดลพหุมิติระหว่างข้อสอบหรือโมเดล BimFit-51 มาตรวจสอบความตรงข้ามกลุ่มแทนที่จะเป็นโมเดลพหุมิติภายในข้อสอบทั้งโมเดล WimFit-someSt8-51 และโมเดล WimFit-allSt8-51 เนื่องจากทั้งสองโมเดลนี้มีค่าสหสัมพันธ์ความสามารถนักเรียนรายสาระการเรียนรู้ต่ำและติดลบ (ดูตารางที่ 4.5) จึงไม่สามารถใช้โมเดลดังกล่าวประมาณค่าความสามารถนักเรียนในสาระที่ 8 ได้ แม้ว่าทั้งสองโมเดลจะสอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดล BimFit-51 แต่ในเมื่อผลวิเคราะห์ขัดแย้งกับธรรมชาติของแบบสอบก็ไม่สามารถใช้โมเดลเหล่านี้ได้เช่นกัน

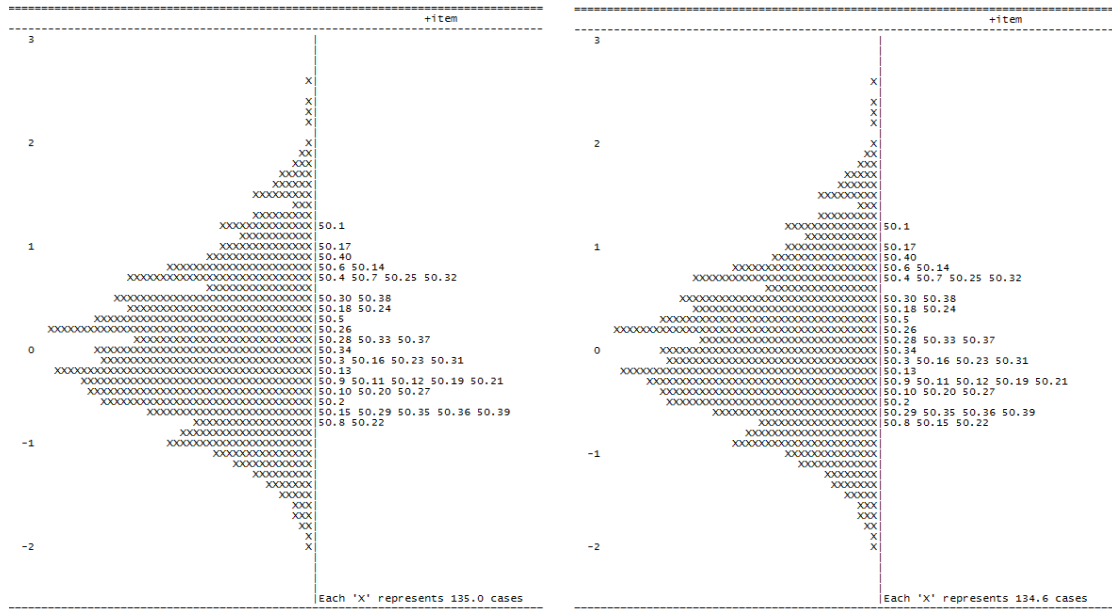
สำหรับการพิจารณาเลือกโมเดลปีการศึกษา 2552 นั้น ผู้วิจัยมีหลักเกณฑ์พิจารณาหลายประการ กล่าวคือ เมื่อให้คะแนนคำตอบแบบสองค่า โมเดลเอกมิติ UniFit-52D เป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด และเมื่อนำโมเดล UniFit-52D มาเปรียบเทียบกับโมเดลเอกมิติ UniFit-52P ที่ให้คะแนนคำตอบแบบหลายค่า พบว่ามีค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันมาก แต่โมเดล UniFit-52P สามารถบรรยายความสามารถของนักเรียนได้ดีกว่าเพราะตำแหน่งค่าความยากข้อสอบครอบคลุมการแจกแจงค่าความสามารถนักเรียนมากกว่า (ไม่สามารถนำดัชนีความสอดคล้องมาเปรียบเทียบกันได้เพราะเมื่อให้คะแนนคำตอบต่างกัน ก็จะได้ว่าเป็นข้อมูลคนละชุดกัน) รวมทั้งโมเดล UniFit-52P ก็มีวิธีให้คะแนนคำตอบเช่นเดียวกับโมเดลที่สอดคล้องกลมกลืนกับข้อมูลที่ให้คะแนนคำตอบแบบหลายค่ามากที่สุดคือ โมเดลพหุมิติภายในข้อสอบ WimFit-someSt8-52P โมเดลนี้ถือว่าเป็นโมเดลที่ประสบความสำเร็จมากที่สุด เนื่องจากสามารถนำไปใช้ประมาณค่าความสามารถนักเรียนได้ครบทุกสาระการเรียนรู้ตามโครงสร้างที่กำหนดไว้ในหลักสูตรแกนกลางฯ พ.ศ. 2551 รวมทั้งมีคุณลักษณะโมเดลสอดคล้องกับความตั้งใจของผู้สร้างแบบสอบ ONET ปีการศึกษา 2552 อีกด้วย

จากคำอธิบายข้างต้น ผู้วิจัยตรวจสอบความตรงข้ามกลุ่มของโมเดลจำนวนทั้งสิ้นหกโมเดล ได้แก่โมเดล UniFit-50 โมเดล BimFit-50 โมเดล UniFit-51 โมเดล BimFit-51 โมเดล UniFit-52P และโมเดล WimFit-someSt8-52P ด้วยวิธีการตรวจสอบความสอดคล้องกลมกลืนของโมเดลเหล่านี้กับข้อมูลอีกชุดหนึ่ง (validating sample) และกำหนดให้แต่ละโมเดลมีค่าพารามิเตอร์ตั้งต้น (initial value) เท่ากับค่าพารามิเตอร์ที่ประมาณได้จากข้อมูลชุดแรก (fitting sample) ไม่ว่าจะเป็นค่าพารามิเตอร์ความยากข้อสอบ ค่าเฉลี่ยความสามารถประชากรนักเรียน ค่าความแปรปรวนและความแปรปรวนร่วมของความสามารถประชากรนักเรียน รายละเอียดของผลการตรวจสอบความตรงข้ามกลุ่มมีดังนี้

ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2550

ตารางที่ 4.10 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2550

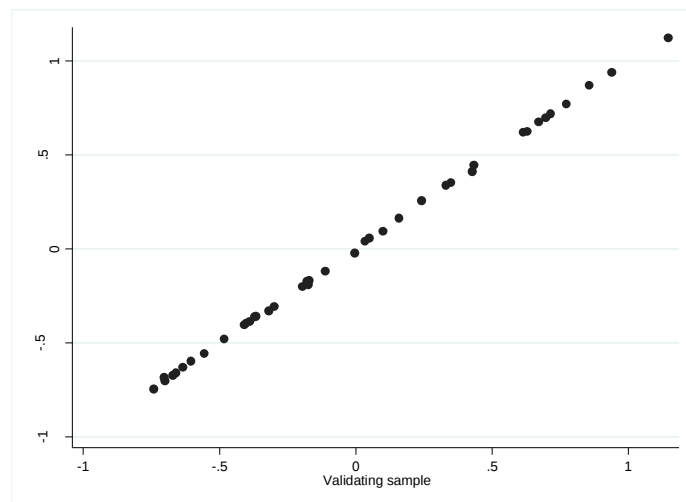
สถิติพื้นฐาน	ค่าประมาณความสามารถนักเรียน		ค่าประมาณความยากข้อสอบ	
	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม
ค่าเฉลี่ย	.007	.009	.000	.000
ความแปรปรวน	.482	.477	.287	.289
ค่าต่ำสุด	-1.778	-1.827	-.747	-.742
ค่าสูงสุด	2.267	2.261	1.120	1.145



(a) Fitting sample

(b) Validating sample

ภาพที่ 4.4 แผนที่ของ Wright ของโมเดลเอกมิติจากการตรวจสอบความตรงข้ามกลุ่ม
ปีการศึกษา 2550



ภาพที่ 4.5 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2550

เมื่อตรวจสอบความสอดคล้องกลมกลืนของโมเดลเอกมิติ UniFit-50 ด้วยข้อมูลคำตอบข้อสอบจากตัวอย่างนักเรียนจำนวน 90,000 คนอีกชุดหนึ่งดังตารางที่ 4.10 พบว่า ค่าประมาณความสามารถนักเรียนจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกันมากกับกลุ่มตัวอย่างความสอดคล้อง และมีค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันมากเท่ากับ .834 และ .835

ตามลำดับ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่ามีความใกล้เคียงกันมากเช่นเดียวกัน รวมทั้งเมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.4 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันมาก แสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล UniFit-50 ด้วยข้อมูลอีกชุดหนึ่ง โมเดล UniFit-50 จะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.11 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิตี ปีการศึกษา 2550

ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง
1	1.120	1.145	.025	21	-.309	-.300	.009
2	-.559	-.556	.003	22	-.705	-.701	.004
3	-.190	-.174	.016	23	-.170	-.170	.000
4	.623	.630	.007	24	.350	.349	.001
5	.256	.241	.015	25	.673	.670	.003
6	.719	.714	.005	26	.163	.158	.005
7	.695	.697	.002	27	-.397	-.402	.005
8	-.747	-.742	.005	28	.092	.099	.007
9	-.361	-.371	.010	29	-.674	-.672	.002
10	-.481	-.482	.001	30	.411	.426	.015
11	-.389	-.390	.001	31	-.174	-.179	.005
12	-.359	-.365	.006	32	.618	.614	.004
13	-.203	-.197	.006	33	.040	.033	.007
14	.768	.773	.005	34	-.022	-.005	.017
15	-.684	-.704	.020	35	-.632	-.635	.003
16	-.120	-.113	.007	36	-.599	-.604	.005
17	.937	.938	.001	37	.056	.049	.007
18	.338	.330	.008	38	.443	.433	.010
19	-.330	-.320	.010	39	-.660	-.660	.000
20	-.405	-.410	.005	40	.869	.856	.013

เมื่อพิจารณาผลต่างของค่าความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้องและกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.11 พบว่าค่าผลต่างอยู่ระหว่าง .001 ถึง .025 ซึ่งถือว่าน้อยมากและเป็นหลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความยากข้อสอบที่ประมาณได้จากโมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันที่สุ่มจากประชากรเดียวกัน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่ำนี้อาจสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.5 และเมื่อตรวจสอบความสัมพันธ์

ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าสูงมากเช่นกัน เท่ากับ .99 แสดงให้เห็นว่าค่าความยากข้อสอบจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีลำดับเช่นเดียวกันกับกลุ่มตัวอย่างความสอดคล้อง ดังนั้นจากหลักฐานเชิงประจักษ์ทั้งหมด ผู้วิจัยจึงสรุปผลว่าโมเดล UniFit-50 มีความตรงข้ามกลุ่ม

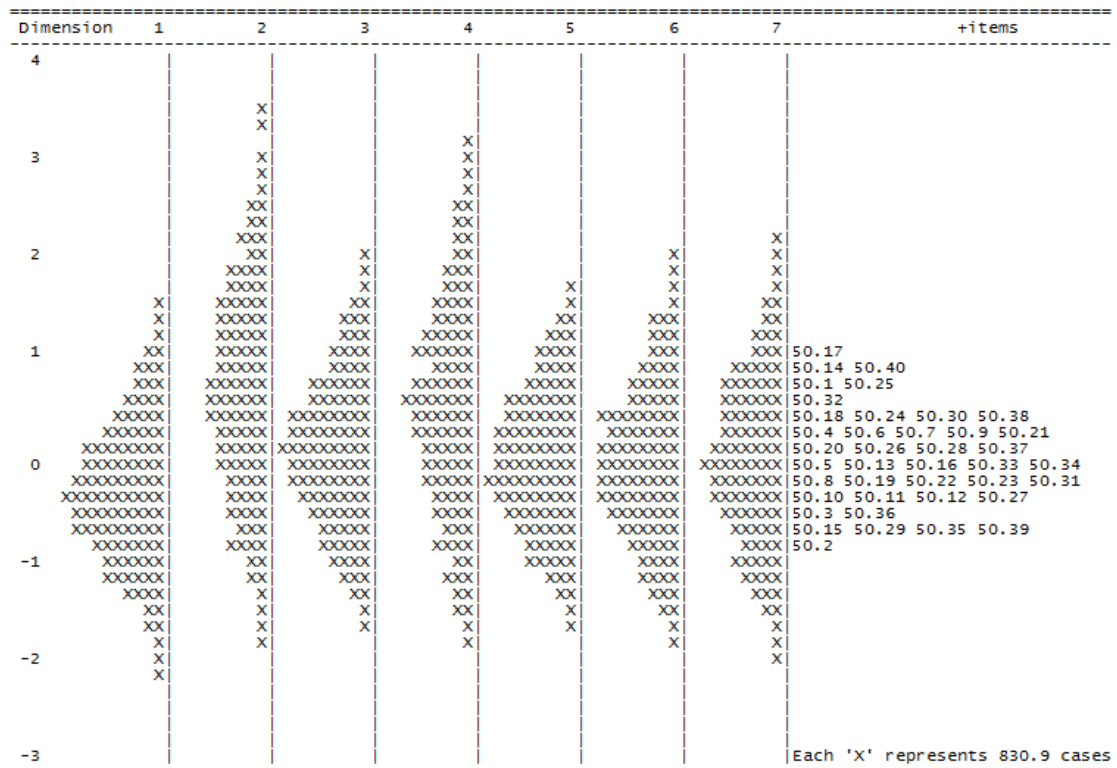
เมื่อตรวจสอบความตรงข้ามกลุ่มของโมเดลพหุมิติระหว่างข้อสอบ BimFit-50 ซึ่งเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดดังตารางที่ 4.12 พบว่าค่าประมาณความสามารถนักเรียนจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกับกลุ่มตัวอย่างความสอดคล้องและมีช่วงค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันระหว่าง .633 ถึง .752 และ .643 ถึง .757 ตามลำดับ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่า มีค่าใกล้เคียงกันมากในทุกสาระการเรียนรู้ และเมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.6 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันมาก แสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล BimFit-50 ด้วยข้อมูลอีกชุดหนึ่ง โมเดล BimFit-50 จะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.12 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลพหุมิติ
ปีการศึกษา 2550

สาระการเรียนรู้	กลุ่มตัวอย่างความสอดคล้อง				กลุ่มตัวอย่างความตรงข้ามกลุ่ม			
	Mean	Variance	Min	Max	Mean	Variance	Min	Max
ค่าความสามารถนักเรียน								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	-.357	.387	-2.288	1.852	-.358	.397	-2.140	1.618
2: ชีวิตกับสิ่งแวดล้อม	.688	1.018	-2.015	3.758	.659	.951	-2.558	3.486
3: สารและสมบัติของสาร	.077	.548	-1.953	2.217	.078	.521	-2.243	2.583
4: แรงและการเคลื่อนที่	.526	.926	-2.131	3.172	.542	.918	-2.657	3.318
5: พลังงาน	-.025	.461	-2.120	1.834	-.036	.456	-2.133	2.053
6: กระบวนการเปลี่ยนแปลงของโลก	-.058	.521	-2.026	2.028	-.047	.534	-2.248	2.541
7: ดาราศาสตร์และอวกาศ	-.011	.719	-2.401	2.728	-.005	.655	-2.505	2.424
ค่าความยากข้อสอบ								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.000	.341	-.938	.749	.000	.345	-.941	.767
2: ชีวิตกับสิ่งแวดล้อม	.000	.099	-.223	.223	.000	.092	-.214	.214
3: สารและสมบัติของสาร	.000	.321	-.648	1.028	.000	.319	-.665	1.021
4: แรงและการเคลื่อนที่	.000	.052	-.255	.186	.000	.052	-.252	.191
5: พลังงาน	.000	.217	-.693	.683	.000	.215	-.687	.678
6: กระบวนการเปลี่ยนแปลงของโลก	.000	.210	-.699	.599	.000	.213	-.706	.592
7: ดาราศาสตร์และอวกาศ	.000	.502	-.742	.905	.000	.482	-.726	.886

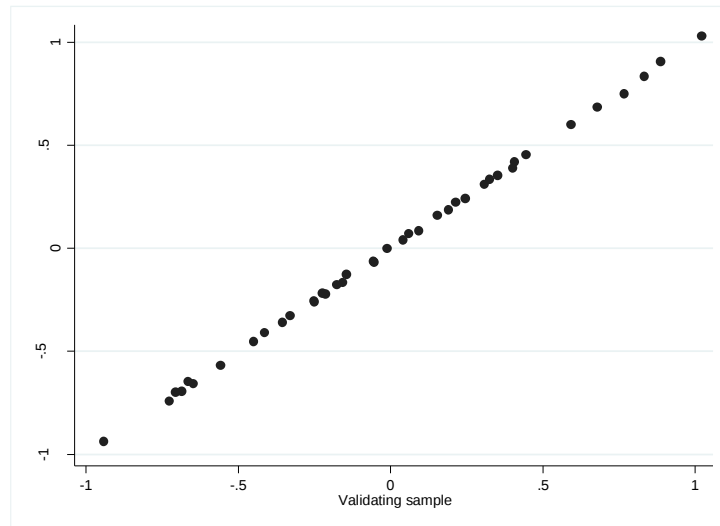


(a) Fitting sample



(b) Validating sample

ภาพที่ 4.6 แผนที่ของ Wright ของโมเดลพหุมิติระหว่างข้อสอบจากการตรวจสอบความตรงข้ามกลุ่ม
ปีการศึกษา 2550



ภาพที่ 4.7 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิติระหว่างข้อสอบ
ปีการศึกษา 2550

ตารางที่ 4.13 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติ ปีการศึกษา 2550

ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง
1	.749	.767	.018	21	.186	.191	.005
2	-.938	-.941	.003	22	-.255	-.252	.003
3	-.570	-.559	.011	23	-.178	-.177	.001
4	.242	.245	.003	24	.353	.352	.001
5	-.127	-.145	.018	25	.683	.678	.005
6	.335	.326	.009	26	.160	.154	.006
7	.310	.308	.002	27	-.410	-.413	.003
8	-.223	-.214	.009	28	.085	.093	.008
9	.223	.214	.009	29	-.693	-.687	.006
10	-.453	-.449	.004	30	.388	.402	.014
11	-.359	-.355	.004	31	-.218	-.224	.006
12	-.328	-.330	.002	32	.599	.592	.007
13	-.166	-.157	.009	33	-.001	-.011	.010
14	.833	.833	.000	34	-.068	-.053	.015
15	-.648	-.665	.017	35	-.699	-.706	.007
16	-.063	-.056	.007	36	-.657	-.647	.010
17	1.028	1.021	.007	37	.041	.042	.001
18	.419	.408	.011	38	.453	.444	.009
19	-.262	-.250	.012	39	-.742	-.726	.016
20	.069	.060	.009	40	.905	.886	.019

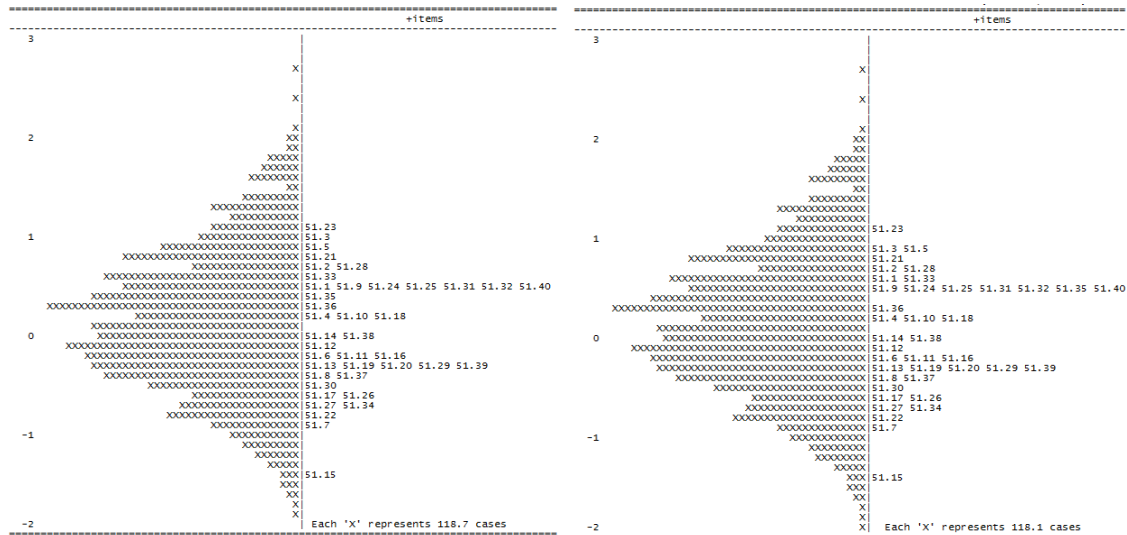
เมื่อพิจารณาผลต่างของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้องและกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.13 พบว่าค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีแนวโน้มเหมือนกัน และมีผลต่างอยู่ระหว่าง .000 ถึง .019 ซึ่งถือว่าน้อยมากและเป็นหลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าความยากข้อสอบที่ประมาณได้จากโมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันซึ่งสุม่จากประชากรเดียวกัน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีค่าสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.7 และเมื่อตรวจสอบความสัมพันธ์ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าสูงมากเช่นกันเท่ากับ .84 แสดงให้เห็นว่าค่าความยากข้อสอบจากกลุ่มตัวอย่างความตรงข้ามกลุ่มส่วนใหญ่มีลำดับเช่นเดียวกันกับกลุ่มตัวอย่างความสอดคล้อง ดังนั้นจากหลักฐานเชิงประจักษ์ทั้งหมด ผู้วิจัยจึงสรุปผลว่าโมเดล BimFit-50 มีความตรงข้ามกลุ่ม

ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2551

เมื่อตรวจสอบความสอดคล้องกลมกลืนของโมเดลเอกมิติ UniFit-51 ซึ่งเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดด้วยข้อมูลคำตอบข้อสอบจากตัวอย่างนักเรียนจำนวน 78,000 คนอีกชุดหนึ่งดังตารางที่ 4.14 พบว่า ค่าประมาณความสามารถนักเรียนจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกันมากกับกลุ่มตัวอย่างความสอดคล้อง และมีค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันมากเท่ากับ .834 และ .833 ตามลำดับ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่ามีค่าใกล้เคียงกันมากเช่นเดียวกันรวมทั้งเมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.8 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันมาก แสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล UniFit-51 ด้วยข้อมูลอีกชุดหนึ่ง โมเดล UniFit-51 จะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.14 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2551

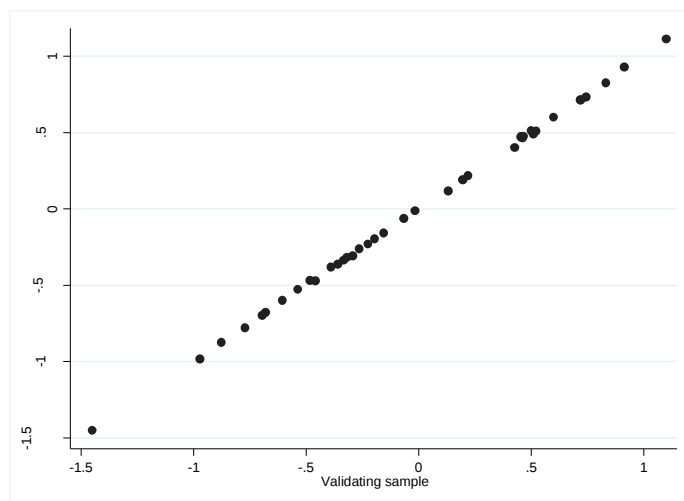
สถิติพื้นฐาน	ค่าประมาณความสามารถนักเรียน		ค่าประมาณความยากข้อสอบ	
	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม
ค่าเฉลี่ย	.113	.116	.000	.000
ความแปรปรวน	.482	.485	.356	.355
ค่าต่ำสุด	-1.764	-1.766	-1.452	-1.451
ค่าสูงสุด	2.337	2.343	1.112	1.099



(a) Fitting sample

(b) Validating sample

ภาพที่ 4.8 แผนที่ของ Wright ของโมเดลเอกมิติจากการตรวจสอบความตรงข้ามกลุ่ม
ปีการศึกษา 2551



ภาพที่ 4.9 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2551

เมื่อพิจารณาผลต่างของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้อง และกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.15 พบว่าค่าความยากข้อสอบจากทั้งสองกลุ่ม ตัวอย่างมีแนวโน้มเหมือนกันและมีผลต่างอยู่ระหว่าง .000 ถึง .026 ซึ่งถือว่าน้อยมากและเป็น หลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความยากข้อสอบที่ประมาณได้จาก โมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันที่สุ่มจากประชากรเดียวกัน นอกจากนี้

สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีค่าสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.9 และเมื่อตรวจสอบความสัมพันธ์ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าสูงมากเช่นกันเท่ากับ .80 แสดงว่าค่าความยากข้อสอบส่วนใหญ่จากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีลำดับเช่นเดียวกันกับกลุ่มตัวอย่างความสอดคล้อง ดังนั้นจากหลักฐานเชิงประจักษ์ทั้งหมด ผู้วิจัยจึงสรุปผลว่าโมเดล UniFit-51 มีความตรงข้ามกลุ่ม

ตารางที่ 4.15 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2551

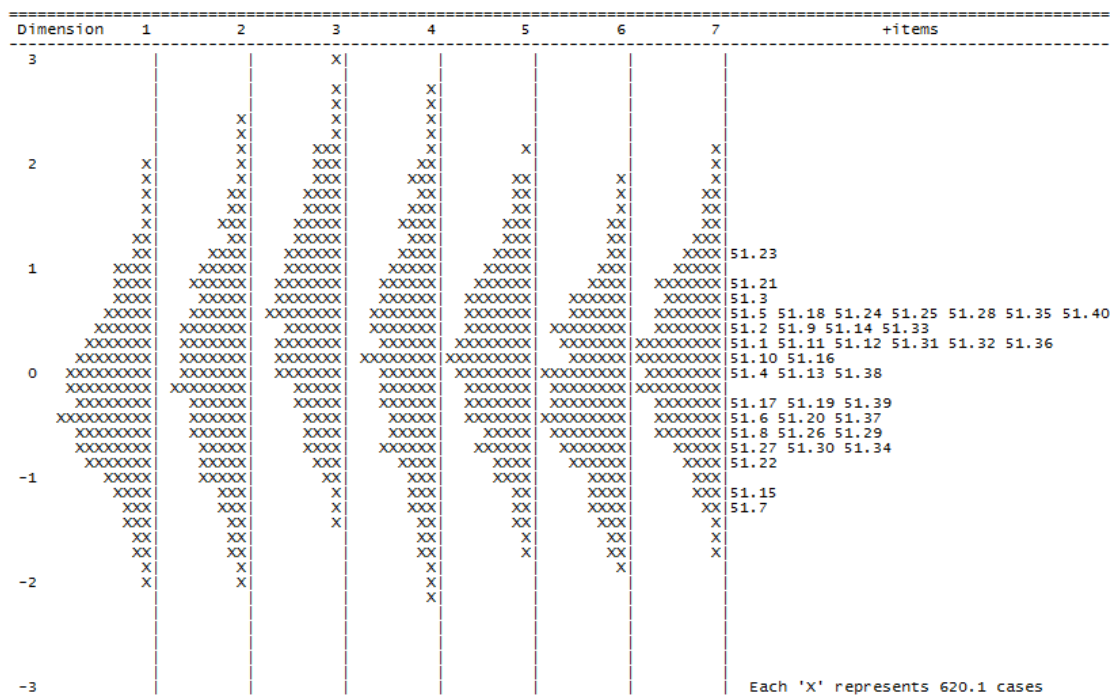
ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง
1	.511	.522	.011	21	.733	.744	.011
2	.716	.718	.002	22	-.873	-.877	.004
3	.930	.913	.017	23	1.112	1.099	.013
4	.189	.196	.007	24	.490	.509	.019
5	.827	.832	.005	25	.475	.464	.011
6	-.230	-.225	.005	26	-.599	-.605	.006
7	-.982	-.973	.009	27	-.697	-.697	.000
8	-.467	-.483	.016	28	.714	.719	.005
9	.474	.456	.018	29	-.318	-.319	.001
10	.192	.196	.004	30	-.529	-.537	.008
11	-.196	-.196	.000	31	.464	.461	.003
12	-.156	-.156	.000	32	.470	.454	.016
13	-.382	-.389	.007	33	.601	.600	.001
14	-.012	-.016	.004	34	-.779	-.771	.008
15	-1.452	-1.451	.001	35	.401	.427	.026
16	-.260	-.265	.005	36	.220	.220	.000
17	-.679	-.682	.003	37	-.470	-.460	.010
18	.118	.130	.012	38	-.063	-.067	.004
19	-.336	-.334	.002	39	-.307	-.293	.014
20	-.363	-.361	.002	40	.512	.499	.013

เมื่อตรวจสอบความตรงข้ามกลุ่มของโมเดลพหุมิติระหว่างข้อสอบ BimFit-51 ดังตารางที่ 4.16 พบว่าค่าประมาณความสามารถนักเรียนจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกับกลุ่มตัวอย่างความสอดคล้อง และมีช่วงค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันระหว่าง .618 ถึง .755 และ .601 ถึง .748 ตามลำดับ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่า มีค่าใกล้เคียงกันมากในทุกสาระการเรียนรู้ และเมื่อพิจารณา

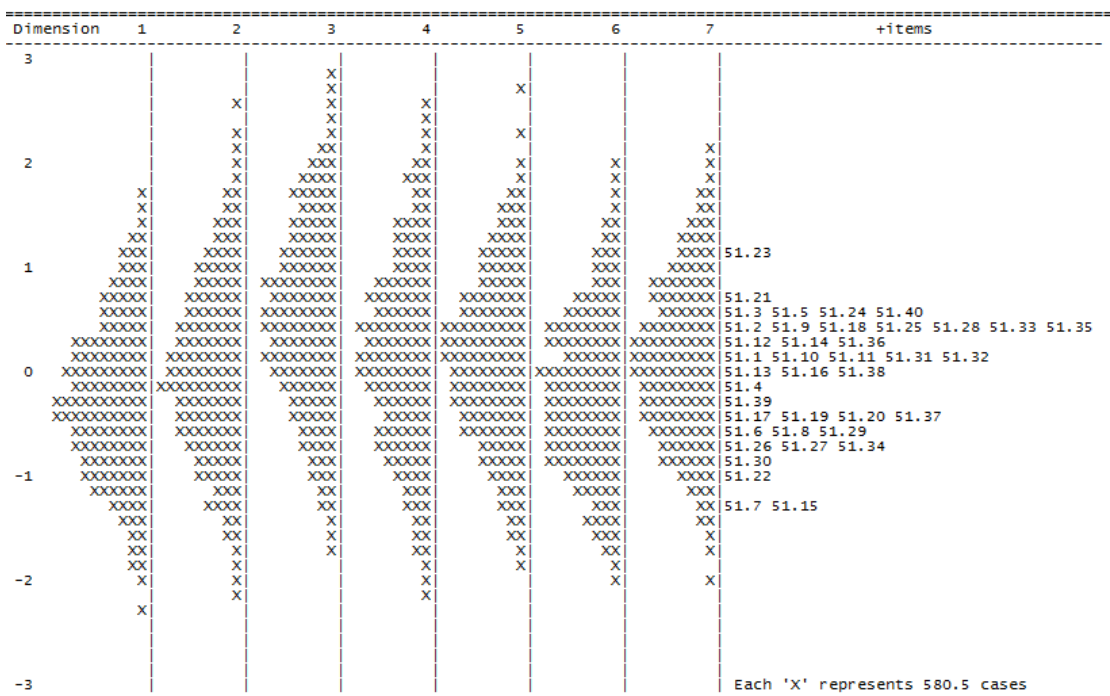
แผนที่ของ Wright ในภาพที่ 4.10 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันมาก แสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล BimFit-51 ด้วยข้อมูลอีกชุดหนึ่ง โมเดลดังกล่าวจะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.16 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลพหุมิติ
ปีการศึกษา 2551

สาระการเรียนรู้	กลุ่มตัวอย่างความสอดคล้อง				กลุ่มตัวอย่างความตรงข้ามกลุ่ม			
	Mean	Variance	Min	Max	Mean	Variance	Min	Max
ค่าความสามารถนักเรียน								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	-.205	.483	-2.251	2.264	-.204	.471	-2.254	2.133
2: ชีวิตกับสิ่งแวดล้อม	.044	.626	-2.323	2.399	.070	.566	-2.354	2.329
3: สารและสมบัติของสาร	.573	.723	-2.149	2.909	.572	.723	-2.188	2.934
4: แรงและการเคลื่อนที่	.174	.726	-2.516	2.851	.176	.683	-2.401	2.735
5: พลังงาน	.151	.518	-2.260	2.432	.155	.534	-2.157	2.552
6: กระบวนการเปลี่ยนแปลงของโลก	-.142	.526	-2.469	2.255	-.135	.531	-2.619	2.244
7: ดาราศาสตร์และอวกาศ	.136	.522	-2.114	2.314	.131	.504	-2.220	2.137
ค่าความยากข้อสอบ								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.000	.505	-1.306	.674	.000	.495	-1.297	.650
2: ชีวิตกับสิ่งแวดล้อม	.000	.266	-.568	.439	.000	.258	-.564	.423
3: สารและสมบัติของสาร	.000	.293	-1.176	.526	.000	.292	-1.170	.535
4: แรงและการเคลื่อนที่	.000	.472	-.411	.793	.000	.466	-.409	.788
5: พลังงาน	.000	.700	-.903	1.152	.000	.700	-.905	1.142
6: กระบวนการเปลี่ยนแปลงของโลก	.000	.295	-.793	.477	.000	.297	-.796	.486
7: ดาราศาสตร์และอวกาศ	.000	.240	-.726	.594	.000	.236	-.725	.573



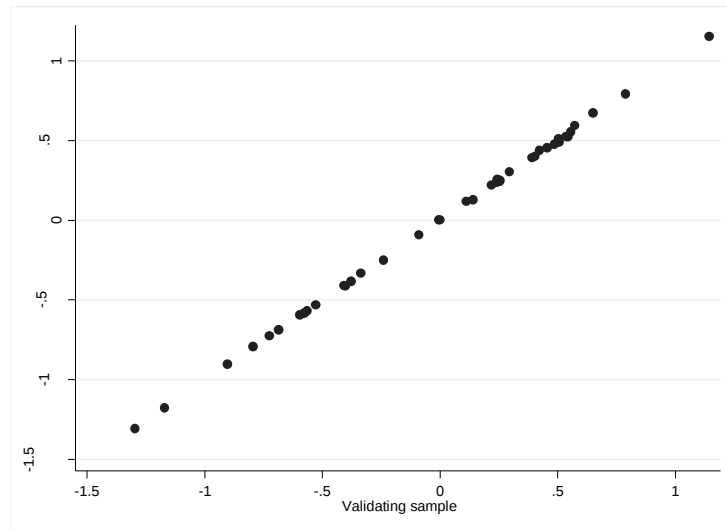
(a) Fitting sample



(b) Validating sample

ภาพที่ 4.10 แผนที่ของ Wright ของโมเดลพหุมิติระหว่างข้อสอบจากการตรวจสอบ

ความตรงข้ามกลุ่ม ปีการศึกษา 2551



ภาพที่ 4.11 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิติระหว่างข้อสอบ
ปีการศึกษา 2551

ตารางที่ 4.17 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติ ปีการศึกษา 2551

ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง
1	.246	.254	.008	21	.793	.788	.005
2	.456	.454	.002	22	-.903	-.905	.002
3	.674	.650	.024	23	1.152	1.142	.010
4	-.093	-.089	.004	24	.521	.543	.022
5	.556	.556	.000	25	.511	.503	.008
6	-.532	-.529	.003	26	-.593	-.597	.004
7	-1.306	-1.297	.009	27	-.688	-.686	.002
8	-.568	-.564	.004	28	.477	.486	.009
9	.439	.423	.016	29	-.581	-.578	.003
10	.128	.141	.013	30	-.793	-.796	.003
11	.219	.218	.001	31	.240	.241	.001
12	.252	.255	.003	32	.255	.243	.012
13	.001	-.001	.002	33	.400	.404	.004
14	.392	.391	.001	34	-.726	-.725	.001
15	-1.176	-1.170	.006	35	.489	.507	.018
16	.118	.110	.008	36	.302	.295	.007
17	-.333	-.337	.004	37	-.412	-.403	.009
18	.526	.535	.009	38	.003	-.006	.009
19	-.382	-.379	.003	39	-.250	-.240	.010
20	-.411	-.409	.002	40	.594	.573	.021

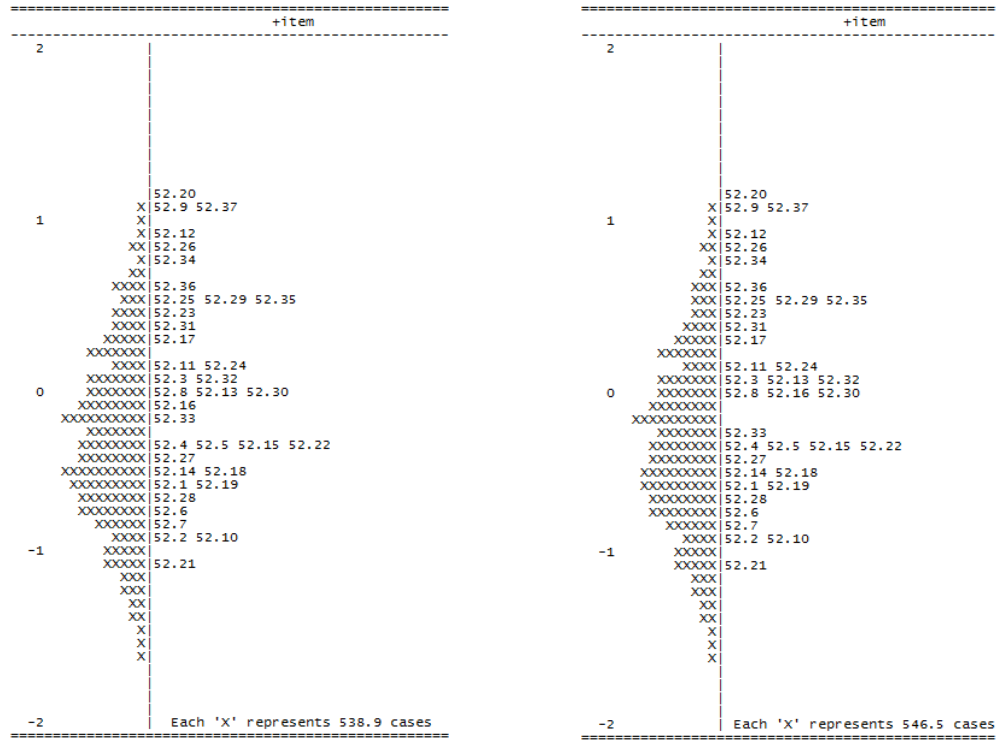
เมื่อพิจารณาผลต่างของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้องและกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.17 พบว่าค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีแนวโน้มเหมือนกันและมีผลต่างอยู่ระหว่าง .000 ถึง .024 ซึ่งถือว่าน้อยมากและเป็นหลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความยากข้อสอบที่ประมาณได้จากโมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันซึ่งสุมจากประชากรเดียวกัน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีค่าสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.11 และเมื่อตรวจสอบความสัมพันธ์ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าสูงมากเช่นกันเท่ากับ .81 แสดงให้เห็นว่าค่าความยากข้อสอบจากกลุ่มตัวอย่างความตรงข้ามกลุ่มส่วนใหญ่มีลำดับเช่นเดียวกันกับกลุ่มตัวอย่างความสอดคล้อง ดังนั้นจากหลักฐานเชิงประจักษ์ทั้งหมด ผู้วิจัยจึงสรุปผลว่าโมเดล BimFit-51 มีความตรงข้ามกลุ่ม

ความตรงข้ามกลุ่มของโมเดลที่สอดคล้องกลมกลืนมากที่สุด ปีการศึกษา 2552

เมื่อตรวจสอบความสอดคล้องกลมกลืนของโมเดลเอกมิติ UniFit-52P ด้วยข้อมูลคำตอบข้อสอบจากตัวอย่างนักเรียนจำนวน 88,000 คนอีกชุดหนึ่งดังตารางที่ 4.18 พบว่า ค่าประมาณความสามารถนักเรียนจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกับกลุ่มตัวอย่างความสอดคล้องและมีค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันมากเท่ากับ .758 และ .755 ตามลำดับ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่ามีค่าใกล้เคียงเช่นเดียวกัน รวมทั้งเมื่อพิจารณาแผนที่ของ Wright ในภาพที่ 4.12 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันแสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล UniFit-52P ด้วยข้อมูลอีกชุดหนึ่ง โมเดลดังกล่าวจะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.18 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2552

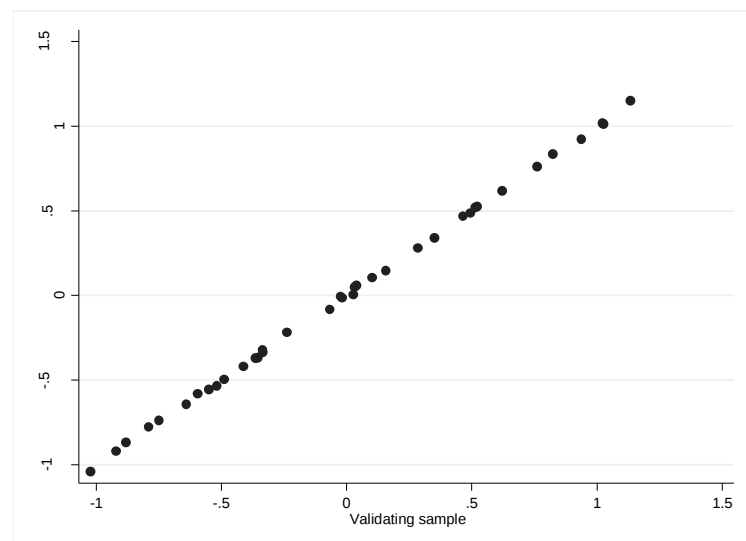
สถิติพื้นฐาน	ค่าประมาณความสามารถนักเรียน		ค่าประมาณความยากข้อสอบ	
	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม
ค่าเฉลี่ย	-.304	-.307	.070	.071
ความแปรปรวน	.257	.262	.411	.408
ค่าสูงสุด	-1.684	-1.653	-1.040	-1.023
ค่าต่ำสุด	1.571	1.613	1.151	1.133



(a) Fitting sample

(b) Validating sample

ภาพที่ 4.12 แผนที่ของ Wright ของโมเดลเอกมิติจากการตรวจสอบความตรงข้ามกลุ่ม
ปีการศึกษา 2552



ภาพที่ 4.13 แผนภาพการกระจายของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2552

เมื่อพิจารณาผลต่างของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้อง และกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.19 พบว่าค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีแนวโน้มเหมือนกัน และมีผลต่างอยู่ระหว่าง .001 ถึง .023 ซึ่งถือว่าน้อยมากและเป็นหลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความยากข้อสอบที่ประมาณได้จากโมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันที่สุ่มจากประชากรเดียวกัน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีค่าสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.13 และเมื่อตรวจสอบความสัมพันธ์ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าปานกลางเท่ากับ .64 แสดงว่าค่าความยากข้อสอบจากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีลำดับค่อนข้างสลับกันไปมากกับกลุ่มตัวอย่างความสอดคล้อง แต่ถ้าพิจารณาลำดับค่าความยากในภาพรวมจะพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ Spearman rank เท่ากับ .99 ดังนั้นจากหลักฐานเชิงประจักษ์ทั้งหมด ผู้วิจัยจึงสรุปผลว่าโมเดล UniFit-52P มีความตรงข้ามกลุ่ม

ตารางที่ 4.19 ผลต่างของค่าความยากข้อสอบจากโมเดลเอกมิติ ปีการศึกษา 2552

ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง ความสอดคล้อง	กลุ่มตัวอย่าง ความตรงข้ามกลุ่ม	ผลต่าง
1	-.555	-.551	.004	21	-1.040	-1.023	.017
2	-.870	-.883	.013	22	-.370	-.364	.006
3	.049	.032	.017	23	.468	.466	.002
4	-.337	-.334	.003	24	.103	.102	.001
5	-.323	-.336	.013	25	.525	.522	.003
6	-.739	-.750	.011	26	.836	.824	.012
7	-.779	-.790	.011	27	-.420	-.413	.007
8	-.007	-.024	.017	28	-.644	-.640	.004
9	1.019	1.023	.004	29	.485	.493	.008
10	-.921	-.922	.001	30	-.014	-.019	.005
11	.146	.157	.011	31	.341	.351	.010
12	.924	.938	.014	32	.059	.039	.020
13	.004	.027	.023	33	-.219	-.239	.020
14	-.534	-.518	.016	34	.762	.761	.001
15	-.367	-.354	.013	35	.521	.514	.007
16	-.083	-.067	.016	36	.619	.621	.002
17	.280	.284	.004	37	1.014	1.026	.012
18	-.498	-.489	.009				
19	-.583	-.596	.013				
20	1.151	1.133	.018				

เมื่อตรวจสอบความตรงข้ามกลุ่มของโมเดลพหุมิติภายในข้อสอบ WimFit-52P ดังตารางที่ 4.20 พบว่าค่าประมาณความสามารถนักเรียนทุกสาระการเรียนรู้จากกลุ่มตัวอย่างความตรงข้ามกลุ่มมีค่าสถิติพื้นฐานใกล้เคียงกับกลุ่มตัวอย่างความสอดคล้องยกเว้นสาระที่ 1 และมีช่วงค่าความเที่ยงแบบ EAP/PV ใกล้เคียงกันระหว่าง .286 ถึง .653 และ .271 ถึง .653 ตามลำดับ โดยสาระที่ 8 มีค่าความเที่ยงน้อยกว่าสาระอื่นๆ อย่างเห็นได้ชัดไม่ว่าจะมาจากกลุ่มตัวอย่างใดก็ตาม เมื่อพิจารณาค่าสถิติพื้นฐานของค่าความยากข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองกลุ่มพบว่า มีค่าใกล้เคียงกันมากในทุกสาระการเรียนรู้ และเมื่อพิจารณาแผนผังของ Wright ในภาพที่ 4.14 พบว่าการแจกแจงค่าความสามารถนักเรียนและค่าความยากข้อสอบมีลักษณะคล้ายคลึงกันมาก แสดงให้เห็นว่าเมื่อตรวจสอบความสอดคล้องของโมเดล WimFit-52P ด้วยข้อมูลอีกชุดหนึ่ง โมเดลดังกล่าวจะให้ค่าประมาณพารามิเตอร์เชิงโครงสร้างไม่แปรเปลี่ยนตามกลุ่มตัวอย่าง

ตารางที่ 4.20 ค่าสถิติพื้นฐานของค่าความสามารถนักเรียนและค่าความยากข้อสอบจากโมเดลพหุมิติ
ปีการศึกษา 2552

สาระการเรียนรู้	กลุ่มตัวอย่างความสอดคล้อง				กลุ่มตัวอย่างความตรงข้ามกลุ่ม			
	Mean	Variance	Min	Max	Mean	Variance	Min	Max
ค่าความสามารถนักเรียน								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.000	.277	-1.610	1.736	.002	.051	-.637	.921
2: ชีวิตกับสิ่งแวดล้อม	.000	.279	-1.434	1.985	.001	.287	-1.462	2.076
3: สารและสมบัติของสาร	.000	.318	-1.731	1.882	.000	.315	-1.727	1.899
4: แรงและการเคลื่อนที่	.000	.263	-1.536	1.935	.000	.264	-1.502	1.951
5: พลังงาน	.000	.279	-1.422	2.261	.000	.292	-1.456	2.303
6: กระบวนการเปลี่ยนแปลงของโลก	.000	.326	-1.673	1.885	.001	.346	-1.700	1.970
7: ดาราศาสตร์และอวกาศ	.001	.146	-1.227	1.294	.001	.153	-1.237	1.359
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	.002	.051	-.637	.921	.003	.058	-.666	.985
ค่าความยากข้อสอบ								
1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	.256	.468	-.590	1.327	.256	.472	-.595	1.347
2: ชีวิตกับสิ่งแวดล้อม	-.094	.066	-.275	.087	-.097	.056	-.264	.070
3: สารและสมบัติของสาร	.419	.365	-.332	1.371	.421	.366	-.335	1.373
4: แรงและการเคลื่อนที่	.223	.367	-.264	1.079	.229	.365	-.267	1.083
5: พลังงาน	.271	.820	-.726	1.529	.274	.805	-.707	1.520
6: กระบวนการเปลี่ยนแปลงของโลก	.514	.078	.284	.826	.518	.080	.284	.832
7: ดาราศาสตร์และอวกาศ	.450	.249	-.364	.899	.455	.252	-.356	.908
8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	-.011	.045	-.332	.254	-.009	.046	-.335	.258

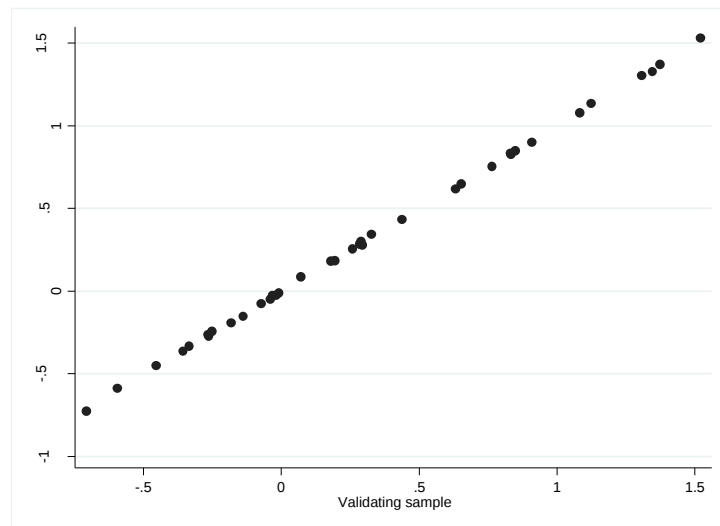
Dimension	1	2	3	4	5	6	7	8	+item
2									
		X				X			
	X	X	X			X			52.20
	X	X	X			X			
		X	XX	X	X	X			52.9 52.12 52.37
	X	X	X	X	X	X			
1	XX	XX	X	X	XX	XX	X		52.26 52.34
	XX	XX	XX	XX	XX	XX	X	X	
	XX	XXX	XX	XX	XXX	XXX	XX	X	52.36
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X	52.23 52.25 52.29 52.35
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	X	52.17
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.31
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.24
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.32
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.8 52.11 52.16 52.30
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.3 52.13
0	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.33
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.4 52.5 52.15 52.22 52.27
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.14 52.18
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.1 52.7 52.19
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.10 52.28
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.6
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.2
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	52.21
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	
-1	XX	XX	XX	XX	XX	XX	XX	X	
	XX	XX	XX	X	X	XX	X		
	X	X	X	X	XX	XX			
	X	X	X	X	X	XX			
-2									Each 'x' represents 920.8 cases

(a) Fitting sample

Dimension	1	2	3	4	5	6	7	8	+item
2									
		X				X			
		X	X			X			
	X	X	X		X	X			52.20
		X	XX		X	X			52.9 52.37
	X	XX	X	XX	X	X			52.12
1	XX	XX	X	XX	XX	XX	X		52.26
	XX	XX	XX	XX	XX	XX	XX	X	52.34
	XXX	XXX	XX	XX	XX	XX	XX	X	52.36
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.23 52.25 52.35
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XX	52.29
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	52.17 52.31
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.24
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.8 52.11 52.16 52.30 52.32
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.3 52.13
0	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.33
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.4 52.5 52.15 52.22 52.27
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.14 52.18
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.1 52.7 52.19
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.10 52.28
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.6
	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	52.2
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	52.21
	XXX	XXX	XXX	XXX	XXX	XXX	XXX	XXX	
-1	XX	XX	XX	XX	XX	XX	XX	X	
	XX	XX	XX	XX	XX	XX	XX		
	X	X	X	X	XX	XX	X		
	X	X	X	X	X	XX			
-2									Each 'x' represents 886.6 cases

(b) Validating sample

ภาพที่ 4.14 แผนที่ของ Wright ของโมเดลพหุมิติภายในข้อสอบจากการตรวจสอบความตรงข้ามกลุ่ม
ปีการศึกษา 2552



ภาพที่ 4.15 แผนภาพการกระจายค่าความยากข้อสอบจากโมเดลพหุมิติภายในข้อสอบ
ปีการศึกษา 2552

ตารางที่ 4.21 ผลต่างของค่าความยากข้อสอบจากโมเดลพหุมิติปีการศึกษา 2552

ข้อที่	กลุ่มตัวอย่าง	กลุ่มตัวอย่าง	ผลต่าง	ข้อที่	กลุ่มตัวอย่าง	กลุ่มตัวอย่าง	ผลต่าง
1	-.275	-.264	.011	21	-.726	-.707	.019
2	-.590	-.595	.005	22	-.051	-.041	.010
3	.182	.179	.003	23	.826	.832	.006
4	-.025	-.020	.005	24	.432	.437	.005
5	-.028	-.033	.005	25	.832	.831	.001
6	-.451	-.454	.003	26	1.134	1.123	.011
7	-.264	-.267	.003	27	-.078	-.074	.004
8	.299	.289	.010	28	-.364	-.356	.008
9	1.371	1.373	.002	29	.755	.765	.010
10	-.332	-.335	.003	30	.284	.284	.000
11	.254	.258	.004	31	.618	.632	.014
12	1.304	1.307	.003	32	.343	.327	.016
13	.185	.194	.009	33	.087	.070	.017
14	-.194	-.183	.011	34	1.079	1.083	.004
15	-.012	-.008	.004	35	.848	.848	.000
16	.279	.293	.014	36	.899	.908	.009
17	.648	.652	.004	37	1.327	1.347	.020
18	-.154	-.139	.015				
19	-.244	-.252	.008				
20	1.529	1.520	.009				

เมื่อพิจารณาผลต่างของค่าประมาณความยากข้อสอบที่ได้จากกลุ่มตัวอย่างความสอดคล้องและกลุ่มตัวอย่างความตรงข้ามกลุ่มในตารางที่ 4.21 พบว่าค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีแนวโน้มเหมือนกัน และมีผลต่างอยู่ระหว่าง .000 ถึง .020 ซึ่งถือว่าน้อยมากและเป็นหลักฐานเชิงประจักษ์บ่งชี้ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความยากข้อสอบที่ประมาณได้จากโมเดลการวัดเดียวกันแต่ใช้ข้อมูลจากกลุ่มตัวอย่างต่างชุดกันซึ่งสุ่มจากประชากรเดียวกัน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ Pearson ของค่าความยากข้อสอบจากทั้งสองกลุ่มตัวอย่างมีค่าสูงมากเท่ากับ .99 และเป็นไปในทิศทางเดียวกันดังภาพที่ 4.15 และเมื่อตรวจสอบความสัมพันธ์ของลำดับค่าความยากข้อสอบด้วยสัมประสิทธิ์สหสัมพันธ์ Spearman rank ก็มีค่าเท่ากับ .71 แสดงว่าค่าความยากข้อสอบจากกลุ่มตัวอย่างความตรงข้ามกลุ่มส่วนใหญ่มีลำดับเช่นเดียวกับกลุ่มตัวอย่างความสอดคล้อง ดังนั้นผู้วิจัยจึงสรุปได้ว่าโมเดล WimFit-52P มีความตรงข้ามกลุ่ม

ตอนที่ 3 สเกลคะแนนความสามารถทางวิทยาศาสตร์ คะแนนจุดตัดขั้นต้นและผลกระทบ

เนื้อหาในตอนนี้เน้นว่าเป็นหัวใจสำคัญของโครงการวิจัย ผลวิเคราะห์ต่างๆ จะเชื่อมโยงเข้าด้วยกันและส่งต่ออย่างเป็นระบบเริ่มตั้งแต่ตอนที่ 2 ของบทที่ 3 เกี่ยวกับการพัฒนาและตรวจสอบแผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์ของนักเรียนเพื่อนำมาใช้ตีความคะแนนจุดตัดขั้นต้นในตอนนี โมเดลการวัดที่ผู้วิจัยพัฒนาขึ้นถูกตรวจสอบคุณภาพอย่างเข้มข้นด้วยวิธีการเชิงปริมาณและเชิงคุณภาพในตอน 1 และ 2 ของบทนี้เพื่อให้มั่นใจได้ว่า โมเดลที่เลือกใช้นั้นถูกต้อง เหมาะสมและน่าเชื่อถือสำหรับพัฒนาต่อไปเป็นสเกลความสามารถนักเรียน เมื่อได้สเกลความสามารถนักเรียนแล้วก็จะเชื่อมต่อเข้ากับแผนที่ภาวะสันนิษฐานเชิงประจักษ์ แล้วจึงกำหนดเป็นคะแนนจุดตัดขั้นต้นในท้ายสุด

ข้อสรุปที่ได้จากผลวิเคราะห์ในตอนที่แล้วคือ โมเดลทั้งหกโมเดลมีความตรงข้ามกลุ่ม สามารถนำไปใช้ประมาณค่าพารามิเตอร์เชิงโครงสร้างได้อย่างถูกต้องและคงเส้นคงวา และมีโมเดลจำนวนครึ่งหนึ่งเป็นโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดด้วย โมเดลทั้งหกนี้ถูกนำมาใช้ประมาณความสามารถนักเรียนสำหรับสร้างเป็นสเกลคะแนนความสามารถตามทฤษฎีการวัดของราล์ซ และนำโมเดลดังกล่าวมาใช้ประมาณค่าความยากข้อสอบสำหรับกำหนดคะแนนจุดตัดขั้นต้น โมเดลเอกมิติ UniFit-50, UniFit-51, และ UniFit-52P จะนำไปใช้สำหรับสร้างสเกลคะแนนความสามารถนักเรียนในภาพรวมและกำหนดคะแนนจุดตัดขั้นต้นในภาพรวมสำหรับแบบสอบ ONET ปีการศึกษา 2550 ถึง 2552 ตามลำดับ ส่วนโมเดลพหุมิติ BimFit-50, BimFit-51, และ WimFit-52P จะถูกนำไปใช้สำหรับสร้างสเกลคะแนนความสามารถนักเรียนรายสาระการเรียนรู้และกำหนดคะแนนจุดตัดขั้นต้นรายสาระการเรียนรู้ให้กับแบบสอบ ONET ปีการศึกษา 2550 ถึง 2552 ตามลำดับ วัตถุประสงค์ของการสร้างสเกลคะแนนคือ นอกจากจะนำไปใช้ตีความคะแนนสอบแล้ว ยังใช้เพื่อแปลงค่าติดลบของค่าประมาณความสามารถให้เป็นค่าบวก ทำให้ผู้ใช้คะแนน ONET ไม่สับสน และยังช่วยกำจัดจุด

ทศนิยมของค่าประมาณความสามารถให้เป็นจำนวนเต็มด้วย วิธีการสร้างสเกลคะแนนความสามารถของโครงการวิจัยนี้คือ เนื่องจากกลุ่มตัวอย่างสำหรับตรวจสอบความสอดคล้องกลมกลืนและกลุ่มตัวอย่างสำหรับตรวจสอบความตรงข้ามกลุ่มให้ผลวิเคราะห์ไม่แตกต่างกัน จึงสามารถรวมกันได้เพื่อประมาณค่าความสามารถนักเรียนสำหรับสร้างเป็นสเกลความสามารถ และนำค่าประมาณความยากข้อสอบมาใช้กำหนดคะแนนจุดตัดขั้นต้น ดังนั้นกลุ่มตัวอย่างนักเรียนสำหรับสร้างสเกลคะแนนความสามารถและกำหนดคะแนนจุดตัดขั้นต้นจึงมีขนาดเท่ากับ 180,000 คนในปีการศึกษา 2550 156,000 คนในปีการศึกษา 2551 และ 176,600 คนในปีการศึกษา 2552

เนื้อหาสาระในตอนนี้แบ่งออกเป็นสามส่วน ส่วนแรกเกี่ยวข้องกับสเกลความสามารถนักเรียนทั้งแบบภาพรวมและรายละเอียดการเรียนรู้ในแต่ละปีการศึกษา ส่วนที่สองเป็นคะแนนจุดตัดขั้นต้นที่ผู้วิจัยเสนอแนะบนพื้นฐานของข้อมูลเชิงประจักษ์ และส่วนสุดท้ายเป็นผลวิเคราะห์ผลกระทบของการนำคะแนนจุดตัดขั้นต้นไปใช้ รายละเอียดของเนื้อหาแต่ละส่วนมีดังนี้

สเกลคะแนนความสามารถทางวิทยาศาสตร์

สเกลคะแนนความสามารถในภาพรวม

ค่าความสามารถนักเรียน (θ) ที่ประมาณได้จากโมเดลเอกมิติในแต่ละปีการศึกษาจะถูกนำมาแปลงค่าเป็นคะแนนสเกล (scaled score: θ^*) ที่มีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 ตามสมการ $\theta^* = 10(\theta) + 50$ เมื่อพิจารณาค่าสถิติพื้นฐานของค่าความสามารถในตารางที่ 4.22 พบว่าความสามารถนักเรียนในแต่ละปีการศึกษาแตกต่างกัน และไม่สามารถกล่าวได้ว่านักเรียนปีการศึกษา 2551 เก่งกว่านักเรียนปีการศึกษา 2550 และ 2552 เนื่องจากนักเรียนในแต่ละปีการศึกษามาจากต่างประชากรกัน รวมทั้งแบบสอบ ONET ของแต่ละปีก็ไม่สามารถกล่าวได้ว่าเป็นแบบสอบคู่ขนาน (parallel test) หรือแบบสอบสมมูล (equivalent test) ดังนั้นผู้วิจัยจึงสร้างสเกลความสามารถนักเรียนแยกรายปีการศึกษาดังแสดงในภาพที่ 4.16

ตารางที่ 4.22 ค่าสถิติพื้นฐานค่าประมาณความสามารถในภาพรวม

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	50.10	.01	52.23	.11	38.70	-.31
ส่วนเบี่ยงเบนมาตรฐาน	17.68	.69	17.51	.70	14.73	.51
ค่าต่ำสุด	2.50	-1.83	2.50	-1.77	2.50	-1.69
ค่าสูงสุด	100.00	2.26	100.00	2.34	100.00	1.61

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
100	2.26	73	100	2.34	73			
		72	97.5	2.21	72			
97.5	2.14	71	95	2.07	71			
95	2.01	70			70			
92.5	1.87	69	92.5	1.92	69			
		68	90	1.78	68			
90	1.73	67			67			
87.5	1.60	66	87.5	1.64	66	97.5-100	1.58-1.61	66
85	1.47	65	85	1.50	65	92.5-96	1.49-1.54	65
		64	82.5	1.37	64	88.5-92	1.36-1.43	64
82.5	1.34	63			63	86-88	1.28-1.33	63
80	1.21	62	80	1.24	62	83.5-85.5	1.19-1.23	62
77.5	1.08	61	77.5	1.11	61	81-83	1.09-1.13	61
75	.97	60	75	1.00	60	78.5-80.5	1.00-1.02	60
72.5	.86	59	72.5	.89	59	76-78	.89-.93	59
70	.75	58	70	.78	58	73.5-75.5	.79-.83	58
67.5	.65	57	67.5	.68	57	71-73	.68-.74	57
65	.55	56	65	.58	56	67.5-70.5	.55-.65	56
62.5	.45	55	62.5	.48	55	65-67	.46-.54	55
60	.36	54	60	.38	54	62-64.5	.36-.45	54
57.5	.26	53	57.5	.29	53	59.5-61.5	.27-.32	53
55	.17	52	55	.20	52	57-59	.17-.23	52
52.5	.08	51	52.5	.11	51	54.5-56.5	.09-.14	51
50	-.01	50	50	.01	50	51.5-54	-.04-.05	50
47.5	-.10	49	47.5	-.08	49	47.5-51	-.14- -.05	49
45	-.19	48	45	-.17	48	45-47	-.23- -.18	48
42.5	-.28	47	40-42.5	-.35- -.26	47	42-44.5	-.35- -.26	47
40	-.37	46	37.5	-.44	46	39.5-41.5	-.43- -.37	46
35-37.5	-.55 - -.46	45	35	-.53	45	35.5-39	-.54- -.45	45
32.5	-.64	44	32.5	-.63	44	32.5-35	-.64- -.56	44
30	-.74	43	30	-.73	43	30-32	-.73- -.69	43
27.5	-.85	42	27.5	-.84	42	27.5-29.5	-.83- -.78	42
		41			41	25-27	-.93- -.87	41
25	-.96	40	25	-.96	40	22.5-24.5	-1.04- -.97	40
22.5	-1.08	39	22.5	-1.08	39	20.5-22	-1.08- -1.07	39
20	-1.20	38	20	-1.19	38	18-20	-1.18- -1.16	38
17.5	-1.32	37	17.5	-1.30	37	15.5-17.5	-1.28- -1.27	37
15	-1.43	36	15	-1.41	36	13-15	-1.37- -1.36	36
12.5	-1.53	35	12.5	-1.50	35	8.5-12.5	-1.53- -1.45	35
10	-1.62	34	10	-1.58	34	4.5-7.5	-1.65- -1.59	34
7.5	-1.71	33	5, 7.5	-1.72- -1.66	33	2.5	-1.69	33
2.5- 5	-1.77- -1.83	32	2.5	-1.77	32			

ภาพที่ 4.16 สเกลคะแนนความสามารถทางวิทยาศาสตร์แบบภาพรวม

จากภาพที่ 4.16 สเกลคะแนนดิบของปีการศึกษา 2550 และ 2551 เป็นสเกลคะแนนดั้งเดิมที่เกิดขึ้นจากวิธีให้คะแนนคำตอบแบบสอบค่าของ สทศ. คะแนนดิบทั้งสองปีนี้จึงเกิดจากผลรวมของจำนวนข้อสอบที่นักเรียนตอบถูกคูณด้วยน้ำหนักคะแนนข้อละ 2.5 เพื่อให้มีคะแนนเต็มเท่ากับ 100 คะแนน แม้ว่าคะแนนต่ำสุดที่เป็นไปได้คือ 0 คะแนน แต่ในข้อมูลที่ใช้สร้างสเกลคะแนนความสามารถนี้มีคะแนนดิบต่ำสุด 2.5 คะแนน ดังนั้นเมื่อประมาณค่าความสามารถนักเรียนด้วยโมเดลเอกมิติและแปลงค่าประมาณความสามารถเป็นคะแนนสเกลจึงมั่นใจได้ว่านักเรียนที่ตอบข้อสอบไม่ถูกเลยจะมีคะแนนสเกลต่ำกว่า 32 คะแนนเพียงเล็กน้อยหรืออาจจะเท่ากับ 32 คะแนนก็เป็นได้

ส่วนสเกลคะแนนดิบของปีการศึกษา 2552 นั้นเกิดจากวิธีให้คะแนนคำตอบแบบหลายค่าสำหรับข้อสอบข้อที่ 33 ถึง 37 ซึ่งเมื่อนักเรียนตอบถูกหนึ่งในสองตัวเลือกจะได้คะแนนครึ่งหนึ่งคือ 2 คะแนน ส่วนข้อที่ 1 ถึง 32 มีวิธีให้คะแนนเช่นเดียวกับ สทศ. ดังนั้นช่วงคะแนนดิบที่เป็นไปได้ของ ปีการศึกษา 2552 จึงอยู่ระหว่าง 0 ถึง 100 คะแนนเช่นเดียวกับสเกลคะแนนดิบดั้งเดิม แต่จะมีคะแนนดิบต่ำสุดเท่ากับ 2.5 คะแนน

เมื่อแปลงค่าประมาณความสามารถ เป็นคะแนนสเกลแบบจำนวนเต็ม (rounded scaled score) พบว่าคะแนนสเกลมีช่วงห่างเท่ากันทีละหนึ่งหน่วยหรือหนึ่งคะแนน ช่วงคะแนนสเกลของปีการศึกษา 2550 และ 2551 ซึ่งเท่ากับ 32 ถึง 73 กว้างกว่าของปีการศึกษา 2552 เมื่อพิจารณาสเกลคะแนนความสามารถของแต่ละปีการศึกษาพบว่า มีข้อสังเกตที่น่าสนใจหลายประการ กล่าวคือ คะแนนสเกลที่เพิ่มขึ้นทีละหนึ่งหน่วยมีค่าไม่เท่ากับหนึ่งหน่วยคะแนนดิบ ตัวอย่างเช่นปีการศึกษา 2550 คะแนนสเกลที่เพิ่มขึ้นจาก 45 เป็น 46 (1 หน่วย) มีค่าเท่ากับคะแนนดิบระหว่าง 35 ถึง 40 คะแนน (5 หน่วยหรือตอบถูก 2 ข้อ) ประการที่สอง นักเรียนที่มีคะแนนดิบไม่เท่ากันอาจจะมีคะแนนสเกลเท่ากันได้ เช่น นักเรียนที่ตอบถูกหนึ่งข้อเท่ากับ 2.5 คะแนน กับนักเรียนที่ตอบถูกสองข้อ 5 คะแนน มีความสามารถใกล้เคียงกันเนื่องจากมีคะแนนสเกลเท่ากัน ประการที่สาม มีคะแนนสเกลบางค่าที่ไม่ตรงกับคะแนนดิบได้เลย เช่น คะแนนสเกลที่เท่ากับ 70 จะไม่ตรงกับคะแนนดิบได้เลย เนื่องจากวิธีให้คะแนนคำตอบเป็นแบบเพิ่มขึ้นทีละ 2.5 คะแนนจึงไม่มีคะแนนดิบที่อยู่ระหว่าง 92.5 ถึง 95 คะแนน แต่ถ้าพิจารณาจากจำนวนข้อสอบที่นักเรียนตอบถูกพบว่าเป็นค่าต่อเนื่องคือตอบถูก 37 ข้อและตอบถูก 38 ข้อ และเมื่อเปรียบเทียบกับวิธีให้คะแนนคำตอบแบบหลายค่าของปีการศึกษา 2552 พบว่าคะแนนสเกลทุกค่าจะสัมพันธ์กับคะแนนดิบที่เกิดขึ้นจริง ประการที่สี่ ค่าสูงสุดและค่าต่ำสุดของคะแนนสเกลแต่ละปีไม่เท่ากัน นักเรียนที่ตอบข้อสอบ ONET ปีการศึกษา 2552 ถูกทุกข้อจะได้คะแนนสเกลเท่ากับ 66 คะแนน แต่นักเรียนที่ตอบข้อสอบ ONET ปีการศึกษา 2550 และ 2551 ถูกทุกข้อจะได้คะแนนสเกลเท่ากันคือ 73 คะแนน ในทางตรงกันข้ามนักเรียนที่ตอบข้อสอบถูกเพียงข้อเดียวในปีการศึกษา 2552 จะได้คะแนนสเกลเท่ากับ 33 คะแนน แต่ถ้าตอบถูกข้อเดียวในปี 2550 และ 2551 กลับได้คะแนนสเกลน้อยกว่าคือ 32 คะแนน ดังนั้นข้อค้นพบเหล่านี้จะช่วยยืนยันว่าไม่สามารถนำคะแนนสเกลมาเปรียบเทียบกับกันระหว่างปีการศึกษาได้

สเกลคะแนนความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

ตารางที่ 4.23 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	7.42	-.36	7.98	-.20	10.62	.50
ส่วนเบี่ยงเบนมาตรฐาน	4.02	.65	4.03	.70	4.69	.80
ค่าต่ำสุด	0	-2.18	0	-2.33	0	-2.05
ค่าสูงสุด	17.50	1.77	17.50	2.32	24.00	2.95

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
		63	17.5	1.25	63	24	1.31	63
		62			62	22	1.21	62
17.5	1.09	61			61			61
		60			60	20-21.5	.96-1.01	60
		59	15	.85	59	19.5	.92	59
		58			58			58
15	.73	57			57	17-19	.65-.75	57
		56			56			56
		55	12.5	.47	55	16.5	.49	55
12.5	.35	54			54	14.5-15	.38-.43	54
		53			53			53
		52			52	12.5-14	.17-.24	52
		51	10	.08	51	12	.13	51
10	-.02	50			50	11.5	-.01	50
		49			49	9.5-10	-.12- -.08	49
		48			48	9	-.22	48
7.5	-.38	47	7.5	-.29	47	7-7.5	-.34- -.33	47
		46			46	6.5	-.43	46
		45			45	4.5-5	-.56- -.55	45
		44			44	4	-.62	44
5	-.71	43	5	-.65	43	2.5	-.77	43
		42			42	2	-.74	42
		41			41			41
2.5	-1.02	40	2.5	-.99	40	0	-.96	40
		39			39			39
		38			38			38
0	-1.30	37	0	-1.33	37			37

ภาพที่ 4.17 สเกลคะแนนความสามารถสาระการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

ค่าสถิติพื้นฐานในตารางที่ 4.23 แสดงให้เห็นว่า ค่าเฉลี่ยค่าประมาณความสามารถนักเรียนปีการศึกษา 2552 แตกต่างจากปีอื่นอย่างชัดเจน เมื่อพิจารณาค่าพิสัยและค่าส่วนเบี่ยงเบนมาตรฐานของค่าประมาณความสามารถปีการศึกษา 2550 ซึ่งน้อยกว่าปีอื่น แสดงให้เห็นว่าความสามารถนักเรียนในปีดังกล่าวแตกต่างกันน้อยกว่าปีการศึกษาที่เหลือ และนักเรียนที่ตอบถูกข้อสอบทุกข้อและตอบผิดทุกข้อที่วัดสาระการเรียนรู้จะมีค่าประมาณความสามารถแตกต่างกัน

เมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.17 พบว่าปีการศึกษา 2552 มีค่าต่ำสุดเท่ากับ 40 คะแนน และมีคะแนนดิบที่เป็นไปได้หลายค่ามากกว่าปีการศึกษา 2550 และ 2551 และจะเห็นว่ามีความถี่หลายค่าที่มีความถี่ใกล้เคียงกันซึ่งแตกต่างจากปีการศึกษา 2550 และ 2551 ที่คะแนนดิบแต่ละค่าจะถูกแปลงเป็นคะแนนสเกลไม่ซ้ำกัน นอกจากนี้ปีการศึกษา 2550 และ 2551 มีจำนวนข้อสอบน้อยกว่าและเมื่อแปลงค่าประมาณความสามารถเป็นคะแนนสเกลแล้วจะมีค่าที่เป็นไปได้ต่ำกว่าปีการศึกษา 2552

สเกลคะแนนความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

เมื่อพิจารณาค่าสถิติพื้นฐานในตารางที่ 4.24 พบว่า ค่าเฉลี่ยค่าประมาณความสามารถนักเรียนแต่ละปีแตกต่างกัน และนักเรียนในปีการศึกษา 2550 และ 2551 มีความสามารถหลากหลายมากกว่าปีการศึกษา 2552 เมื่อพิจารณาการกระจายค่าประมาณความสามารถปีการศึกษา 2550 พบว่าจะกระจายมากกว่าปีอื่นๆ ทั้งนี้เนื่องจากปีการศึกษาดังกล่าวมีข้อสอบเพียงสองข้อที่วัดสาระนี้ ค่าความสามารถที่ประมาณได้จึงมีช่วงกว้างกว่าปีการศึกษา 2552 ที่มีข้อสอบวัดสาระนี้มากกว่าและมีรูปแบบคำตอบที่เป็นไปได้มากกว่า

ตารางที่ 4.24 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	3.13	.66	3.83	.11	5.26	0.00
ส่วนเบี่ยงเบนมาตรฐาน	1.84	.97	2.36	.92	3.38	.53
ค่าต่ำสุด	0	-2.55	0	-2.82	0	-1.57
ค่าสูงสุด	5	3.45	7.50	2.84	10.50	1.73

จำนวนรูปแบบคำตอบที่เป็นไปได้ของสาระการเรียนรู้ทำให้สเกลคะแนนความสามารถทั้งสามปีในภาพที่ 4.18 มีช่วงคะแนนกว้าง นั่นคือค่าความสามารถนักเรียนถูกประมาณได้บางตำแหน่งของสเกล สะท้อนให้เห็นว่าควรจะมีจำนวนข้อสอบมากกว่าสามข้อสำหรับประมาณค่าความสามารถนักเรียนในสาระการเรียนรู้

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
5	1.43	64	7.5	1.25	64	10.5	0.71	64
		63			63			63
		62			62			62
		61			61			61
		60			60			60
		59			59			59
		58			58			58
		57			57			57
		56			56			56
		55			55			55
2.5	.37	54	5	0.41	54	8	0.30	54
		53			53			53
		52			52			52
		51			51			51
		50			50			50
		49			49			49
		48			48			48
		47	2.5	-0.29	47	2.5	-0.29	47
		46			46			46
		45			45			45
0	-.57	44			44	0	-0.60	44
		43			43			43
		42			42			42
		41	0	-0.93	41			41

ภาพที่ 4.18 สเกลคะแนนความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

สเกลคะแนนความสามารถสาระการเรียนรู้สารและสมบัติของสาร

ตารางที่ 4.25 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้สารและสมบัติของสาร

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	12.91	.08	12.19	.50	8.20	0.00
ส่วนเบี่ยงเบนมาตรฐาน	5.57	.74	4.86	.80	4.35	.56
ค่าต่ำสุด	0	-2.30	0	-2.05	0	-1.69
ค่าสูงสุด	25.00	2.71	20.00	2.95	20.00	1.84

จากตารางที่ 4.25 ค่าการกระจายของค่าประมาณความสามารถนักเรียนปีการศึกษา 2550 และ 2551 มีสารสนเทศใกล้เคียงกันสะท้อนให้เห็นว่าความสามารถของนักเรียนในแต่ละปีหลากหลายมากกว่าปีการศึกษา 2552 นอกจากนี้ยังมีค่าเฉลี่ยค่าประมาณความสามารถของแต่ละปีแตกต่างกัน นักเรียนที่ได้คะแนนต่ำสุดและสูงสุดในแต่ละปีการศึกษาจะมีค่าประมาณความสามารถแตกต่างกันด้วย

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded θ^*	คะแนนดิบ	θ	Rounded θ^*	คะแนนดิบ	θ	Rounded θ^*
25	1.74	67	20	1.71	67			67
		66			66			66
		65			65			65
		64			64			64
22.5	1.28	63	17.5	1.28	63	20	1.33	63
		62			62			62
		61			61	17.5	1.07	61
		60			60			60
20	.93	59	15	.87	59			59
		58			58	15	.78	58
		57			57			57
17.5	.61	56			56			56
		55	12.5	.50	55	12.5	.49	55
		54			54			54
15	.32	53			53			53
		52	10	.15	52	10	.19	52
		51			51			51
12.5	.01	50			50			50
		49			49	7.5	-.10	49
		48	7.5	-.19	48			48
10	-.29	47			47			47
		46			46	5	-.37	46
		45	5	-.51	45			45
7.5	-.57	44			44	2.5	-.61	44
		43			43			43
5	-.83	42	2.50	-.78	42			42
		41			41	0	-.86	41
		40	0	-1.01	40			40
2.5	-1.09	39			39			39
		38			38			38
0	-1.33	37			37			37

ภาพที่ 4.19 สเกลคะแนนความสามารถสาระการเรียนรู้สารและสมบัติของสาร

เมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.19 พบว่าในแต่ละปีการศึกษาจะมีคะแนนสเกลความสามารถต่ำสุดและสูงสุดไม่เท่ากัน ในปีการศึกษา 2550 คะแนนดิบแต่ละค่าจะมีตำแหน่งกระจายบนสเกลคะแนนความสามารถค่อนข้างสม่ำเสมอ แสดงให้เห็นว่าข้อสอบปีดังกล่าวสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ดีเมื่อเปรียบเทียบกับปีการศึกษาอื่น นอกจากนี้ในปีการศึกษา 2552 กลุ่มนักเรียนที่มีคะแนนดิบสูงระหว่าง 17.5 ถึง 20 จะมีคะแนนสเกลต่างกันเพียงสองคะแนน แต่ปีการศึกษา 2551 จะต่างกันถึง 4 คะแนน

สเกลคะแนนความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่

ค่าสถิติพื้นฐานในตารางที่ 4.26 แสดงให้เห็นว่า ค่าเฉลี่ยค่าประมาณความสามารถนักเรียนปีการศึกษา 2552 แตกต่างจากปีอื่นอย่างชัดเจน เมื่อพิจารณาค่าพิสัยและค่าส่วนเบี่ยงเบนมาตรฐานของค่าประมาณความสามารถปีการศึกษา 2550 และ 2551 พบว่ามีค่าใกล้เคียงกัน แสดงว่าการกระจายความสามารถนักเรียนในแต่ละปีคล้ายคลึงกัน และแตกต่างอย่างชัดเจนกับปีการศึกษา 2552 นอกจากนี้เมื่อพิจารณาค่าต่ำสุดและค่าสูงสุดของค่าประมาณความสามารถในแต่ละปีพบว่า นักเรียนที่ตอบถูกทุกข้อและตอบผิดทุกข้อจะมีค่าประมาณความสามารถแตกต่างกันในแต่ละปีด้วย

สเกลคะแนนความสามารถดังภาพที่ 4.20 แสดงให้เห็นว่า คะแนนสเกลต่ำสุดของแต่ละปีการศึกษาใกล้เคียงกันแต่ค่าสูงสุดต่างกัน ในปีการศึกษา 2552 มีคะแนนดิบสองค่าคือ 6.5 และ 7 คะแนนที่เมื่อแปลงเป็นคะแนนสเกลแล้วมีค่าเท่ากัน นั่นหมายความว่านักเรียนที่มีคะแนนดิบดังกล่าวมีความสามารถใกล้เคียงกัน นอกจากนี้พิสัยของคะแนนสเกลยังแคบกว่าของปีการศึกษา 2550 และ 2551 ด้วยทั้งนี้เนื่องจากปีการศึกษา 2552 มีข้อสอบวัดสาระนั้นมากกว่าปีอื่นและให้คะแนนคำตอบแบบหลายค่า ทำให้มีรูปแบบคำตอบที่เป็นไปได้มากขึ้นและรูปแบบคำตอบแต่ละแบบจะสัมพันธ์กับความสามารถนักเรียนจึงสามารถประมาณค่าความสามารถได้หลายตำแหน่งมากขึ้นบนสเกลคะแนนความสามารถ สเกลคะแนนความสามารถปีการศึกษา 2552 จึงสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ดีกว่าปีการศึกษา 2550 และ 2551 ซึ่งมีข้อสอบที่วัดสาระนั้นน้อยกว่าและให้คะแนนคำตอบแบบสองค่า

ตารางที่ 4.26 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	4.49	.54	3.98	.37	6.32	0.00
ส่วนเบี่ยงเบนมาตรฐาน	2.47	.98	2.36	.96	3.16	.51
ค่าต่ำสุด	0	-2.73	0	-2.64	0	-1.48
ค่าสูงสุด	7.50	3.34	7.50	3.33	14.00	2.04

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
7.5	1.56	66			66			66
		65	7.5	1.53	65			65
		64			64			64
		63			63			63
		62			62			62
		61			61	14	1.13	61
		60			60			60
		59			59			59
		58			58	12	.75	58
5	.67	57			57	11.5	.65	57
		56	5	.60	56	10	.55	56
		55			55			55
		54			54	9.5	.38	54
		53			53	9	.29	53
		52			52	7.5	.18	52
		51			51			51
		50			50	6.5-7	-.04- .05	50
2.5	-.15	49	2.5	-.10	49	5	-.15	49
		48			48			48
		47			47	4.5	-.27	47
		46			46	4	-.35	46
		45			45	2.5	-.48	45
		44			44	2	-.56	44
		43	0	-.69	43			43
		42			42	0	-.77	42
0	-.86	41			41			41

ภาพที่ 4.20 สเกลคะแนนความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่

สเกลคะแนนความสามารถสาระการเรียนรู้พลังงาน

ตารางที่ 4.27 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้พลังงาน

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	8.62	-.04	7.97	-.05	6.06	0.00
ส่วนเบี่ยงเบนมาตรฐาน	4.12	.68	3.61	.68	3.11	.53
ค่าต่ำสุด	0	-2.15	0	-2.35	0	-1.40
ค่าสูงสุด	17.50	2.09	15.00	1.96	14.00	2.40

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
17.5	1.29	63			63	14	1.29	63
		62			62			62
		61	15	1.06	61			61
		60			60			60
15	.89	59			59	12	.85	59
		58			58	11.5	.79	58
		57			57			57
		56	12.5	.64	56	10	.57	56
12.5	.50	55			55	9.5	.47	55
		54			54	9	.40	54
		53			53			53
		52	10	.24	52	7.5	.20	52
10	.14	51			51	6.5-7	.05-.10	51
		50			50			50
		49	7.5	-.15	49	5	-.13	49
7.5	-.22	48			48	4.5	-.24	48
		47			47	4	-.28	47
		46			46			46
		45	5	-.50	45	2.5	-.46	45
5	-.56	44			44	2	-.56	44
		43			43	0	-.75	43
		42	2.5	-.83	42			42
2.5	-.86	41			41			41
		40			40			40
0	-1.14	39	0	-1.14	39			39

ภาพที่ 4.21 สเกลคะแนนความสามารถสาระการเรียนรู้พลังงาน

จากตารางที่ 4.27 ค่าการกระจายของค่าประมาณความสามารถนักเรียนปีการศึกษา 2550 และ 2551 มีสารสนเทศใกล้เคียงกันสะท้อนให้เห็นว่าความสามารถของนักเรียนหลากหลายมากกว่า ปีการศึกษา 2552 นอกจากนี้ นักเรียนที่ได้คะแนนต่ำสุดและสูงสุดในแต่ละปีการศึกษายังมี ค่าประมาณความสามารถแตกต่างกันด้วย เมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.21 พบว่าค่าต่ำสุดของคะแนนสเกลในแต่ละปีการศึกษาค่อนข้างแตกต่างกัน แต่มีค่าสูงสุดใกล้เคียงกัน แม้ว่าในปีการศึกษา 2550 และ 2551 คะแนนดิบจะมีตำแหน่งกระจายบนสเกลคะแนนความสามารถ ค่อนข้างสม่ำเสมอ แต่พบว่ามีช่องว่างจำนวนมากบนสเกลทำให้สเกลนี้สั้นน้อยกว่าปีการศึกษา 2552 ซึ่งสเกลมีช่องว่างคะแนนน้อยกว่า นั่นหมายความว่าคะแนนสเกลจะสอดคล้องกับคะแนนดิบที่ แตกต่างกันและคะแนนสเกลแต่ละคะแนนก็บ่งชี้ความสามารถที่ไม่เหมือนกันด้วย ลักษณะสเกล คะแนนความสามารถปีการศึกษา 2552 จึงสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ดีกว่า

สเกลคะแนนความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

ตารางที่ 4.28 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	7.31	-.04	7.04	-.02	2.93	0.00
ส่วนเบี่ยงเบนมาตรฐาน	3.77	.73	3.71	.68	2.27	.58
ค่าต่ำสุด	0	-2.20	0	-2.36	0	-1.71
ค่าสูงสุด	15.00	2.51	15.00	2.15	7.50	2.05

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552		
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded
		θ^*			θ^*			θ^*
15	1.34	63			63			63
		62	15	1.16	62			62
		61			61			61
		60			60			60
		59			59	7.5	.92	59
12.5	.84	58	12.5	.78	58			58
		57			57			57
		56			56			56
		55			55			55
10	.38	54	10	.41	54	5	.39	54
		53			53			53
		52			52			52
		51			51			51
		50	7.5	.03	50			50
7.5	-.05	49			49	2.5	-.10	49
		48			48			48
		47	5	-.33	47			47
		46			46			46
5	-.45	45			45	0	-.54	45
		44			44			44
		43	2.5	-.67	43			43
2.5	-.81	42			42			42
		41	0	-.94	41			41
		40			40			40
0	-1.12	39			39			39

ภาพที่ 4.22 สเกลคะแนนความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

จากตารางที่ 4.28 ค่าการกระจายของค่าประมาณความสามารถนักเรียนแต่ละปีการศึกษา แตกต่างกันอย่างชัดเจนให้เห็นว่าความสามารถนักเรียนในแต่ละปีมีรูปแบบหลากหลาย ค่าเฉลี่ย ค่าประมาณความสามารถของทั้งสามปีการศึกษาที่ใกล้เคียงกันแสดงให้เห็นว่าโดยภาพรวมนักเรียนมีแนวโน้มความสามารถอยู่ในระดับใกล้เคียงกัน นอกจากนี้นักเรียนที่ได้คะแนนต่ำสุดและสูงสุดในแต่ละปีการศึกษายังมีค่าประมาณความสามารถแตกต่างกันด้วย

เมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.22 พบว่าค่าต่ำสุดและค่าสูงสุดของคะแนนสเกลปีการศึกษา 2552 ค่อนข้างแตกต่างจากปีการศึกษาอื่น คะแนนดิบแต่ละค่าของแต่ละปี มีตำแหน่งกระจายของสเกลคะแนนความสามารถและมีช่องว่างระหว่างคะแนนจำนวนมากด้วย ลักษณะสเกลแบบนี้แสดงให้เห็นว่าสเกลคะแนนทั้งสามปีการศึกษาสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ปานกลาง

สเกลคะแนนความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ

จากตารางที่ 4.29 ค่าการกระจายของค่าประมาณความสามารถนักเรียนปีการศึกษา 2550 และ 2551 มีสารสนเทศใกล้เคียงกันสะท้อนให้เห็นว่าความสามารถของนักเรียนหลากหลายมากกว่าปีการศึกษา 2552 นอกจากนี้นักเรียนที่ได้คะแนนต่ำสุดและสูงสุดในแต่ละปีการศึกษายังมีค่าประมาณความสามารถแตกต่างกันด้วย

เมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.23 พบว่าค่าต่ำสุดและค่าสูงสุดของปีการศึกษา 2552 แตกต่างอย่างชัดเจนจากปีการศึกษาอื่น ในปีการศึกษา 2552 พบว่ามีคะแนนดิบหลายค่าที่มีเมื่อแปลงเป็นคะแนนสเกลแล้วมีค่าเท่ากัน นั่นหมายความว่านักเรียนที่ได้คะแนนดิบ 9.5 และ 10 คะแนนเป็นกลุ่มนักเรียนที่มีความสามารถเดียวกัน นอกจากนี้การเรียงตัวต่อเนื่องของคะแนนดิบบนสเกลคะแนนความสามารถของปีการศึกษา 2552 แสดงให้เห็นว่าสเกลดังกล่าวสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ดีในช่วงคะแนนสเกล 45 ถึง 58 ซึ่งแตกต่างจากปี 2550 และ 2551 ซึ่งแม้ว่าจะบรรยายขอบเขตความสามารถนักเรียนได้กว้างกว่าปีการศึกษา 2552 แต่มีช่องว่างกว้างและจำนวนมากจึงทำให้บรรยายโครงสร้างความสามารถนักเรียนได้น้อยกว่า

ตารางที่ 4.29 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ

สถิติพื้นฐาน	ปีการศึกษา 2550		ปีการศึกษา 2551		ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	180,000	180,000	156,000	156,000	176,600	176,600
ค่าเฉลี่ย	6.22	-.01	9.23	.09	5.55	0.00
ส่วนเบี่ยงเบนมาตรฐาน	3.37	.83	4.22	.74	2.98	.38
ค่าต่ำสุด	0	-2.55	0	-2.26	0	-1.23
ค่าสูงสุด	12.50	2.47	17.50	2.45	14.00	1.36

(a) ปีการศึกษา 2550			(b) ปีการศึกษา 2551			(c) ปีการศึกษา 2552				
คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded	คะแนนดิบ	θ	Rounded		
		θ^*			θ^*			θ^*		
12.5	1.42	65	17.5	1.47	65			65		
		64			64			64		
		63			63			63		
		62			62			62		
		61			61			61		
10	.81	60	15	.99	60			60		
		59			59			59		
		58			58	14	.82	58		
		57			57	12	.68	57		
		56	12.5	.56	56	56	56			
55	55	11.5			.49	55				
54	54	9.5-10			.41	54				
53	53					53				
7.5	.21	52			10	.16	52	7-9	.16-.23	52
		51	51	51						
		50	50	6.5			0.00	50		
		49	49	4.5-5			-.08 - -.07	49		
		5	-.31	48	7.5	-.20	48	2-2.5	-.31- -.30	47
47	47			47						
46	46							46		
45	5			-.53			45	0	-.54	45
44					44	44				
43		43	43							
42		2.5	-.84		42	42				
41	41			41						
40	40			40						
39	39			39						
0	-1.19	38	0	-1.16	38					38

ภาพที่ 4.23 สเกลคะแนนความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ

สเกลคะแนนความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

ตามที่ได้ผู้วิจัยได้กล่าวไว้ก่อนหน้านี้ เนื่องจากโมเดลที่ใช้ประมาณค่าความสามารถนักเรียน สาระการเรียนรู้ของปีการศึกษา 2550 และ 2551 ไม่สอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดลพหุ มิติระหว่างข้อสอบและโมเดลเอกมิติ จึงไม่เหมาะสมที่จะใช้โมเดลเหล่านั้นประมาณค่าความสามารถ นักเรียนในสาระการเรียนรู้นี้ แต่สำหรับปีการศึกษา 2552 โมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดคือโมเดลพหุมิติภายในข้อสอบ WimFit-someSt8-52p และให้ค่าประมาณความสามารถ นักเรียนในสาระนี้ด้วย ทำให้ค่าสถิติพื้นฐานและสเกลคะแนนความสามารถจึงมีเฉพาะ ปีการศึกษา 2552 เท่านั้น

จากตารางที่ 4.30 ค่าการกระจายของค่าประมาณความสามารถนักเรียนมีค่าน้อยแสดงว่าความสามารถนักเรียนไม่หลากหลายมากนัก และเมื่อพิจารณาสเกลคะแนนความสามารถดังภาพที่ 4.24 พบว่ามีการเรียงตัวค่อนข้างต่อเนื่องของคะแนนดิบบนสเกลคะแนนความสามารถ แสดงว่าสเกลดังกล่าวสามารถบรรยายโครงสร้างความสามารถนักเรียนได้ดีในช่วงคะแนนสเกล 45 ถึง 55

ตารางที่ 4.30 ค่าสถิติพื้นฐานของความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

สถิติพื้นฐาน	ปีการศึกษา 2552	
	คะแนนดิบ	ค่าประมาณ
จำนวนตัวอย่าง	176,600	176,600
ค่าเฉลี่ย	10.06	0.00
ส่วนเบี่ยงเบนมาตรฐาน	4.78	.23
ค่าต่ำสุด	0	-.66
ค่าสูงสุด	20	1.01

คะแนนดิบ	θ	Rounded θ^*
20	.54	55
17.5	.37	54
		53
15	.22	52
12.5	.11	51
10	0.00	50
7.5	-.12	49
5	-.23	48
		47
2.5	-.35	46
0	-.49	45

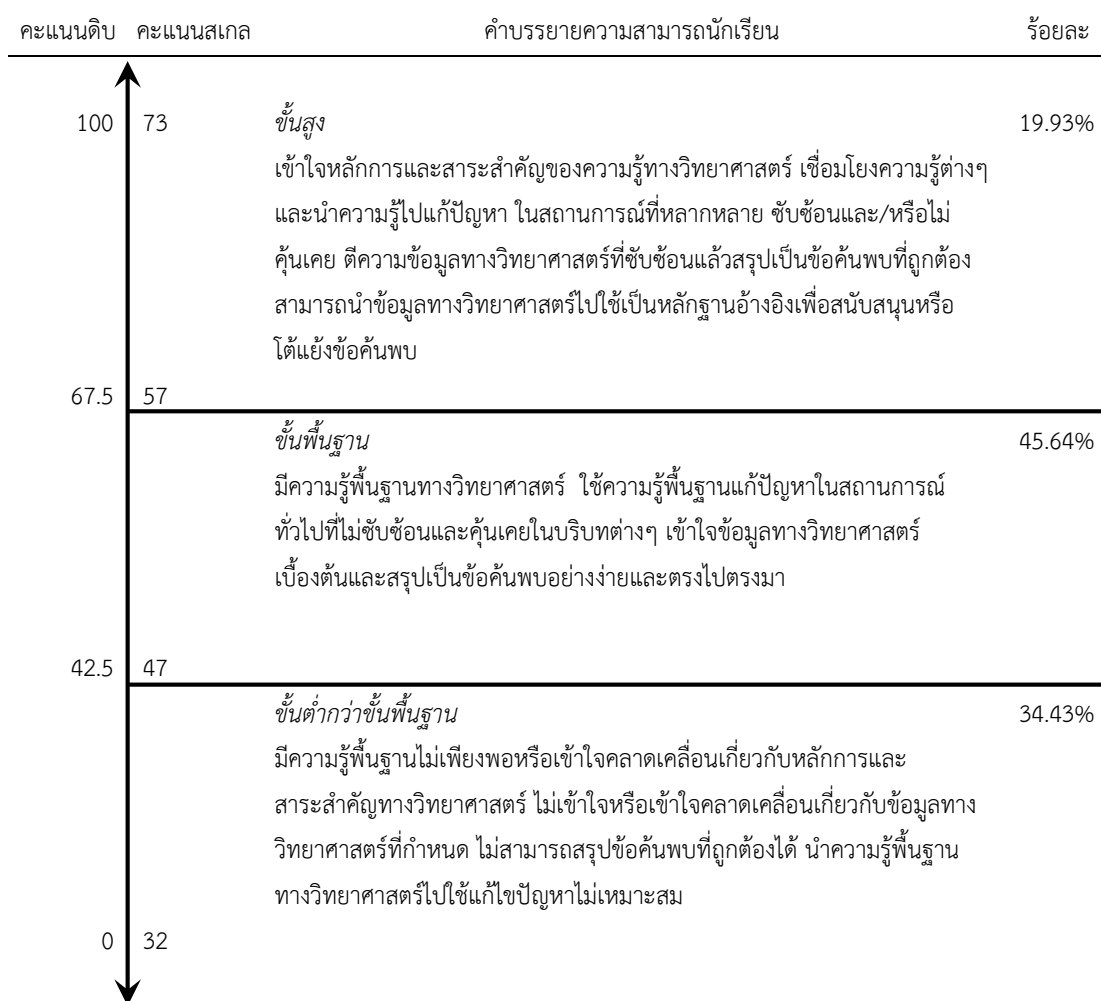
ภาพที่ 4.24 สเกลคะแนนความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

คะแนนจุดตัดขั้นต้นและผลกระทบ

วิธีกำหนดคะแนนจุดตัดขั้นต้นของโครงการวิจัยนี้เป็นการหาค่าเฉลี่ยค่าความยากข้อสอบที่ได้จากโมเดลทั้งหกโมเดลที่ผู้วิจัยได้คัดเลือกไว้แล้วและนำโมเดลเหล่านั้นมาสร้างเป็นสเกลคะแนนเพื่อเชื่อมโยงให้เห็นความสัมพันธ์กันระหว่างคะแนนดิบและคะแนนสเกลที่แปลงมาจากค่าความสามารถ ลูกศรสองหัวที่ชี้ขึ้นบนและลงล่างเป็นแกนกลางสำคัญที่เชื่อมสเกลคะแนนความสามารถเข้ากับแผนที่

คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2550

คะแนนจุดตัดขั้นต้นในภาพรวมคือค่าเฉลี่ยค่าประมาณความยากข้อสอบ ONET ของแต่ละปีการศึกษาโดยแบ่งข้อสอบตามระดับความยากหรือระดับความสามารถนักเรียนที่ปรากฏในแผนที่ภาวะสันนิษฐานเชิงประจักษ์ สำหรับปีการศึกษา 2550 พบว่าค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C1.1 ในภาคผนวก) ระดับขั้นสูงจำนวน 12 ข้อเท่ากับ .67 ลอจิต เมื่อแปลงเป็นคะแนนสเกลเท่ากับ 56.71 หรือ 57 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานจำนวน 28 ข้อ เท่ากับ -.26 ลอจิต เมื่อแปลงเป็นคะแนนสเกลมีค่าเท่ากับ 47.38 หรือ 47 คะแนน นอกจากนี้ในกระบวนการวิเคราะห์คะแนนจุดตัดผู้วิจัยใช้กลุ่มตัวอย่างขนาดใหญ่จึงทำให้ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (standard error of estimation) มีค่าต่ำด้วย เมื่อคำนวณช่วงความเชื่อมั่นที่ระดับ 95% ของค่าเฉลี่ยความยากข้อสอบในแต่ละระดับจึงพบว่ามีช่วงแคบมาก ค่าเฉลี่ยความยากข้อสอบที่คำนวณได้ทั้งสองระดับจึงคงเส้นคงวา



ภาพที่ 4.25 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2550

เมื่อนำคะแนนจุดตัดขั้นต้นที่คำนวณได้ในแต่ละระดับไปจับคู่กับคะแนนสเกลและเชื่อมโยงกับแผนที่ภาวะสันนิษฐานเชิงประจักษ์ (ภาพที่ 3.19) ดังภาพที่ 4.25 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 0 ถึง 40 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 28% ที่จะตอบถูกข้อสอบขั้นสูง

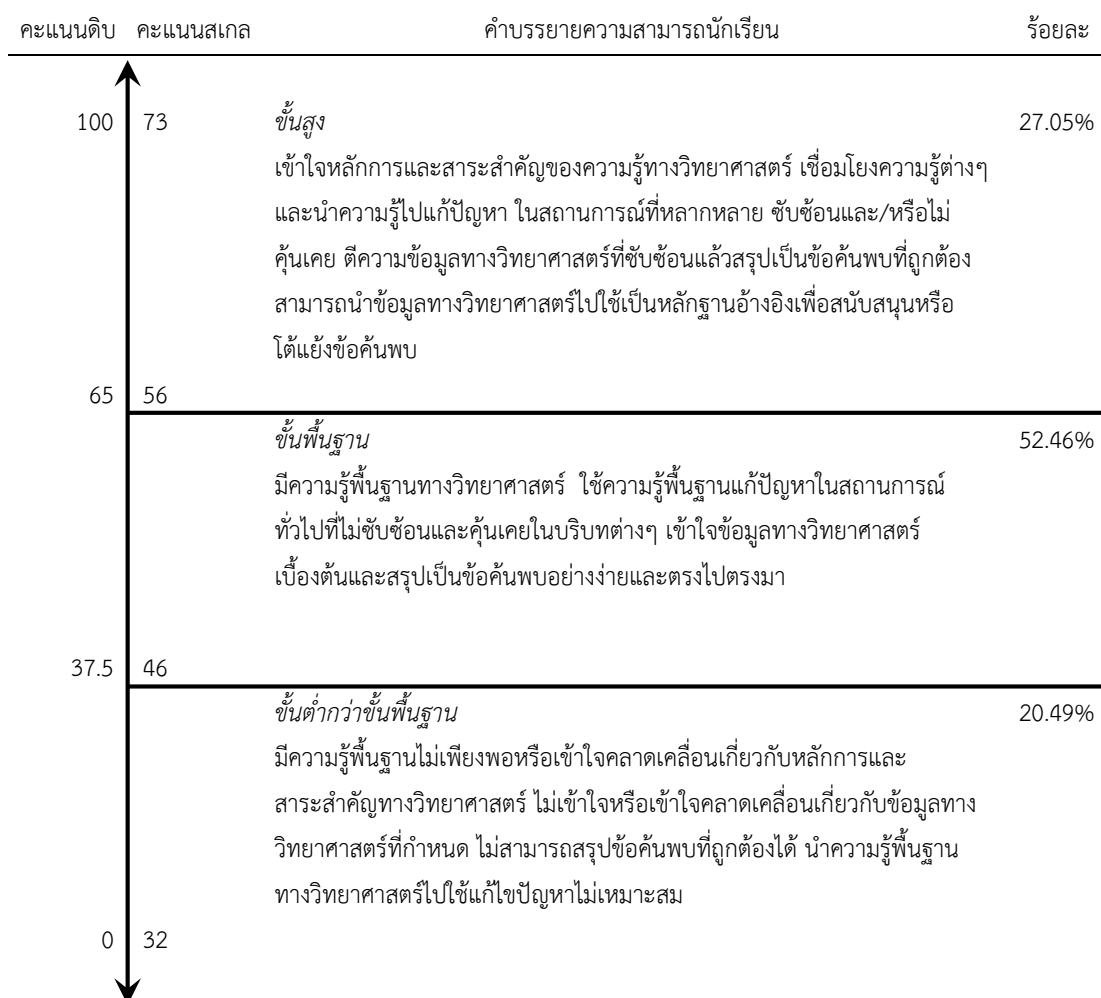
สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบระหว่าง 42.5 ถึง 65 คะแนนหรือมีคะแนนสเกลระหว่าง 47 ถึง 56 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบระหว่าง 67.5 ถึง 100 คะแนนหรือมีคะแนนสเกลระหว่าง 57 ถึง 73 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 72% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นสูง

การตีความคะแนนจุดตัดขั้นต้นของแต่ละระดับความสามารถ สามารถทำได้โดยตรงไปตรงมาและเป็นระบบที่เชื่อมโยงกันระหว่างข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพ ตัวอย่างเช่น นักเรียนที่มีคะแนนดิบต่ำกว่า 42.5 คะแนนจะมีแนวโน้มสูงที่จะมีความรู้พื้นฐานทางวิทยาศาสตร์ไม่เพียงพอหรือเข้าใจคลาดเคลื่อน ยิ่งนักเรียนมีคะแนนต่ำมากเท่าใดก็จะมีแนวโน้มสูงมากขึ้นที่จะมีความสามารถตรงตามคำบรรยายที่กำหนด

จากภาพที่ 4.25 แสดงให้เห็นว่า ถ้าผู้วิจัยมีมุมมองต่อคะแนนจุดตัดขั้นต้นเป็นแบบมาตรฐานความสามารถ (standard) และนำคะแนนจุดตัดทั้งสองค่าไปใช้ประเมินความสามารถนักเรียนจากกลุ่มตัวอย่างจำนวน 180,000 คน จะพบว่า จำนวนนักเรียนส่วนใหญ่มีความสามารถระดับขั้นพื้นฐาน และมีนักเรียนประมาณ 1 ใน 5 คนที่จะมีความสามารถขั้นสูง แต่ถ้าผู้วิจัยมีมุมมองแบบคะแนนผ่านขั้นต่ำสุด (minimum passing score) ก็พบว่าคะแนนดิบ 42.5 คะแนนจะเป็นเพียงจุดตัดเดียวที่ใช้ประเมินความสามารถนักเรียน นั่นหมายความว่าจะมีนักเรียนประมาณ 1 ใน 3 คนที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดนี้

คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2551

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C1.2 ในภาคผนวก) ระดับขั้นสูงจำนวน 16 ข้อเท่ากับ .59 ลอจิตเมื่อแปลงเป็นคะแนนสเกลเท่ากับ 55.85 หรือ 56 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานจำนวน 24 ข้อ เท่ากับ -.39 ลอจิตเมื่อแปลงเป็นคะแนนสเกลมีค่าเท่ากับ 46.10 หรือ 46 คะแนน นอกจากนี้เมื่อคำนวณช่วงความเชื่อมั่นของค่า ความยากข้อสอบที่ระดับ 95% พบว่ามีช่วงแคบมากใกล้เคียงกับค่าเฉลี่ยเนื่องจากใช้กลุ่มตัวอย่างขนาดใหญ่ทำให้ค่าความคลาดเคลื่อนมาตรฐานของการประมาณมีค่าน้อยมาก ดังนั้นจึงสรุปว่าค่าเฉลี่ยความสามารถทั้งสองมีค่าคงเส้นคงวา



ภาพที่ 4.26 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2551

เมื่อนำคะแนนจุดตัดขั้นต้นที่คำนวณได้ของแต่ละระดับจับคู่กับคะแนนสเกลและเชื่อมโยงกับแผนที่ภาวะสันนิษฐานเชิงประจักษ์ (ภาพที่ 3.19 หน้า 92) ดังภาพที่ 4.26 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 0 ถึง 35 คะแนนหรือมีคะแนนสเกลน้อยกว่า 46 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน และมีโอกาสเฉลี่ยอย่างมากที่สุด 27% ที่จะตอบถูกข้อสอบขั้นสูง

สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบระหว่าง 37.5 ถึง 62.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 47 ถึง 55 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบระหว่าง 65 ถึง 100 คะแนนหรือมีคะแนนสเกลตั้งแต่ 57 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 73% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นสูง

สำหรับการตีความคะแนนจุดตัดขั้นต้นนั้น สามารถนำคะแนนดิบหรือคะแนนสเกลของนักเรียนไปเปรียบเทียบกับคะแนนจุดตัดและคำบรรยายระดับความสามารถ นักเรียนที่มีคะแนน

เท่ากับคะแนนจุดตัดขั้นต้นจะถือว่ามีความสามารถตรงกับระดับความสามารถที่สูงกว่า เช่น นักเรียนที่มีคะแนนดิบเท่ากับ 37.5 คะแนนจะพิจารณาให้นักเรียนคนนั้นมีความสามารถอยู่ในขั้นพื้นฐาน และถ้านักเรียนมีคะแนนดิบเท่ากับ 50 คะแนนก็จะยังมีแนวโน้มสูงมากที่ความสามารถจะมากกว่าความยากข้อสอบขั้นพื้นฐาน นั่นหมายความว่าโอกาสที่นักเรียนจะตอบถูกข้อสอบขั้นพื้นฐานจะเพิ่มขึ้นมากกว่า 50% แสดงว่านักเรียนมีความรู้ความเข้าใจเนื้อหาพื้นฐานทางวิทยาศาสตร์และสามารถประยุกต์ใช้ในสถานการณ์ไม่ซับซ้อนได้

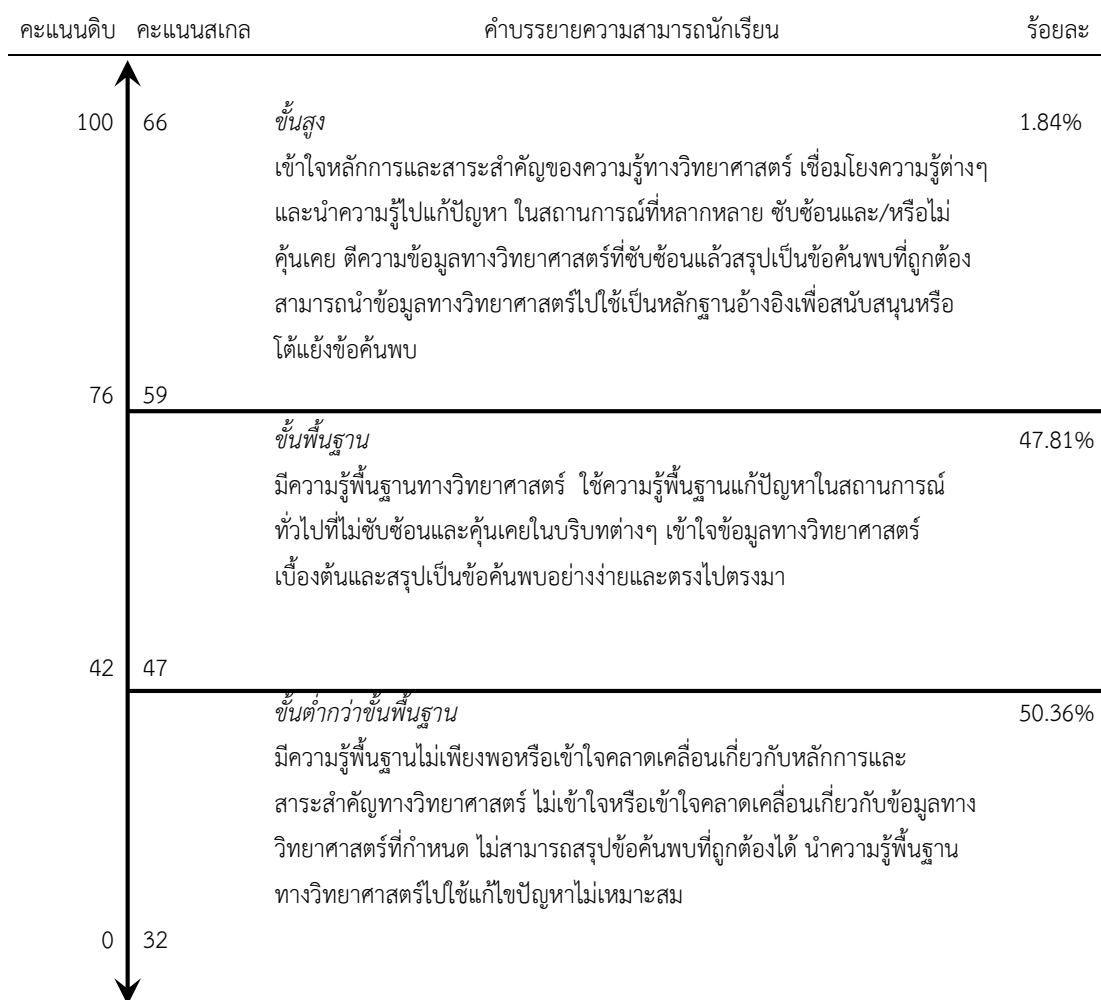
จากภาพที่ 4.26 แสดงให้เห็นว่า ถ้าผู้วิจัยมีมุมมองต่อคะแนนจุดตัดขั้นต้นเป็นแบบมาตรฐานความสามารถและนำคะแนนจุดตัดทั้งสองค่าไปใช้ประเมินความสามารถนักเรียนจากกลุ่มตัวอย่างจำนวน 156,000 คน จะพบว่าจำนวนนักเรียนเกิน 50% มีความสามารถระดับขั้นพื้นฐานและมีนักเรียนประมาณ 1 ใน 4 คนที่จะมีความสามารถขั้นสูง แต่ถ้าผู้วิจัยมีมุมมองแบบคะแนนผ่านขั้นต่ำสุดก็จะพบว่าคะแนนดิบ 37.5 คะแนนจะเป็นเพียงจุดตัดเดียวที่ใช้ประเมินความสามารถนักเรียน นั่นหมายความว่าจะมีนักเรียนประมาณ 1 ใน 5 คนที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดนี้

คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2552

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C1.3 ในภาคผนวก) ระดับขั้นสูงจำนวน 10 ข้อเท่ากับ .91 ลอจิท เมื่อแปลงเป็นคะแนนสเกลเท่ากับ 59.12 หรือ 59 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานจำนวน 27 ข้อ เท่ากับ -.25 ลอจิท เมื่อแปลงเป็นคะแนนสเกลมีค่าเท่ากับ 47.47 หรือ 47 คะแนน นอกจากนี้เมื่อคำนวณช่วงความเชื่อมั่นของค่าความยากข้อสอบที่ระดับ 95% พบว่ามีช่วงแคบมากใกล้กับค่าเฉลี่ย เนื่องจากการวิเคราะห์ในครั้งนี้ใช้กลุ่มตัวอย่างขนาดใหญ่ทำให้ค่าความคลาดเคลื่อนมาตรฐานของการประมาณมีค่าน้อยมาก ค่าเฉลี่ยความสามารถทั้งสองจึงคงเส้นคงวา

เมื่อนำคะแนนจุดตัดขั้นต้นที่คำนวณได้ของแต่ละระดับจับคู่กับคะแนนสเกลและเชื่อมโยงกับแผนที่ภาวะสันนิษฐานเชิงประจักษ์ (ภาพที่ 3.19) ดังภาพที่ 4.27 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 0 ถึง 41.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 24% ที่จะตอบถูกข้อสอบขั้นสูง

สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบระหว่าง 42 ถึง 75.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 47 ถึง 58 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบระหว่าง 76 ถึง 100 คะแนนหรือมีคะแนนสเกลตั้งแต่ 59 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 76% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นสูง



ภาพที่ 4.27 คะแนนจุดตัดขั้นต้นความสามารถทางวิทยาศาสตร์ในภาพรวม ปีการศึกษา 2552

สำหรับการตีความคะแนนจุดตัดขั้นต้นนั้น สามารถนำคะแนนดิบหรือคะแนนสเกลของนักเรียนไปเปรียบเทียบกับคะแนนจุดตัดและคำบรรยายระดับความสามารถ นักเรียนที่มีคะแนนเท่ากับคะแนนจุดตัดขั้นต้นจะถือว่ามีความสามารถตรงกับระดับความสามารถที่สูงกว่า เช่น นักเรียนที่มีคะแนนดิบเท่ากับ 42 คะแนนจะพิจารณาให้นักเรียนคนนั้นมีความสามารถอยู่ในขั้นพื้นฐาน และถ้านักเรียนมีคะแนนดิบเท่ากับ 60 คะแนนก็จะมีแนวโน้มสูงมากๆ ที่ค่าประมาณความสามารถจะมากกว่าค่าความยากข้อสอบขั้นพื้นฐาน และมีแนวโน้มสูงมากที่ค่าประมาณความสามารถจะมากกว่าค่าความยากข้อสอบขั้นสูง นั่นหมายความว่าโอกาสที่นักเรียนจะตอบถูกข้อสอบขั้นพื้นฐานจะเพิ่มขึ้นเรื่อยๆ มากกว่า 50% เมื่อความสามารถนักเรียนสูงขึ้น ความสามารถของนักเรียนก็จะมีแนวโน้มเป็นไปตามคำบรรยายความสามารถตามที่ระบุไว้ในแผนที่ภาวะสันนิษฐานเชิงประจักษ์

จากภาพที่ 4.27 แสดงให้เห็นว่า ถ้าผู้วิจัยมีมุมมองต่อคะแนนจุดตัดขั้นต้นเป็นแบบมาตรฐาน ความสามารถและนำคะแนนจุดตัดทั้งสองค่าไปใช้ประเมินความสามารถนักเรียนจากกลุ่มตัวอย่างจำนวน 176,000 คน จะพบว่าจำนวนนักเรียนเกือบทั้งหมด (98.16%) มีความสามารถต่ำกว่าขั้นสูง แต่ถ้าผู้วิจัยมีมุมมองแบบคะแนนผ่านขั้นต่ำสุดก็จะพบว่าคะแนนดิบ 42 คะแนนจะเป็นเพียงจุดตัด

เดียวที่ใช้ประเมินความสามารถนักเรียน นั้นหมายความว่าจะมีนักเรียนจำนวนครึ่งหนึ่งของทั้งหมดที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดนี้

ข้อค้นพบที่น่าสนใจจากการกำหนดคะแนนจุดตัดขั้นต้นในภาพรวมทั้งสามปีการศึกษา คือแบบสอบ ONET ทั้งสามปีเป็นแบบสอบคนละฉบับ มีข้อกำหนดลักษณะแบบสอบ (test specification) ไม่เหมือนกัน และนักเรียนที่ทำแบบสอบก็มาจากคนละปีการศึกษา ซึ่งโครงการวิจัยนี้พัฒนาโครงสร้างความสามารถนักเรียนที่ใช้ทำแบบสอบในแต่ละปีการศึกษาโดยใช้โครงสร้างความยากข้อสอบที่ข้อสอบทั้งสามปีมีร่วมกัน แล้วเชื่อมโยงไปยังความสามารถนักเรียนที่ใช้ทำข้อสอบในแต่ละส่วนของโครงสร้างความยากข้อสอบ แผนที่ภาวะสันนิษฐานเชิงประจักษ์จึงเป็นเสมือนเครื่องมือที่ใช้เปรียบเทียบและเชื่อมโยงข้อมูลที่มีลักษณะคล้ายคลึงกันในแต่ละส่วนเข้าด้วยกันอย่างเป็นระบบ และเมื่อพิจารณาผลการกำหนดคะแนนจุดตัดขั้นต้นในภาพรวมทั้งสามปีการศึกษาโดยใช้วิธีสร้างแผนที่ความสามารถแฝงเชิงประจักษ์ พบว่า แม้ว่าแบบสอบ ONET ของแต่ละปีการศึกษาจะแตกต่างกันดังกล่าวข้างต้น แต่มีคะแนนจุดตัดที่แปลงให้อยู่ในรูปของคะแนนสเกลใกล้เคียงกัน (แต่มีคะแนนดิบต่างกันอย่างชัดเจน) โดยมีคะแนนจุดตัดแรกที่กั้นระหว่างนักเรียนขั้นต่ำกว่าขั้นพื้นฐานและขั้นพื้นฐานเท่ากับ 47, 46, และ 47 คะแนน ตั้งแต่ปีการศึกษา 2550 ถึง 2552 ตามลำดับ และมีคะแนนจุดตัดที่สองที่กั้นระหว่างนักเรียนขั้นพื้นฐานกับขั้นสูงเท่ากับ 59, 56, และ 57 ตั้งแต่ปีการศึกษา 2550 ถึง 2552 คะแนนตามลำดับ ดังนั้นนักวิจัยจึงมีแนวคิดว่าการกำหนดคะแนนจุดตัดขั้นต้นร่วม (common tentative cutscore) ของแบบสอบ ONET ปีการศึกษา 2550 ถึง 2552 น่าจะสามารถคำนวณได้จากค่าเฉลี่ยของคะแนนจุดตัดขั้นต้นของแต่ละระดับซึ่งก็คือคะแนนสเกลเท่ากับ 47 และ 57 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสำหรับการเรียนรู้สิ่งมีชีวิตและกระบวนการดำรงชีวิต

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ระดับขั้นสูงเท่ากับ .41 ลอจิทหรือคะแนนสเกลเท่ากับ 54 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.55 ลอจิทหรือคะแนนสเกลมีค่าเท่ากับ 45 คะแนน จากภาพที่ 4.28 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 45 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 28% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 7.5 ถึงต่ำกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 45 ถึง 53 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 54 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 72% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความ สามารถ
คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	
17.5	61	18.01%	17.5	63	21.73%	24	63	.82%	↑ ขั้นสูง
12.5	54		12.5	55		22	61		
		40.51%			42.48%			65.01%	↑ ขั้นพื้นฐาน
7.5			7.5						↓ ขั้นต่ำกว่า
	45			44		9	48		
		41.48%			35.79%			34.17%	
0	37		0	37		0	40		

ภาพที่ 4.28 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้สิ่งมีชีวิตและกระบวนการดำรงชีวิต

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.2 ในภาคผนวก) ระดับขั้นสูงเท่ากับ .48 ลอจิทหรือคะแนนสเกลเท่ากับ 55 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.64 ลอจิทหรือคะแนนสเกลเท่ากับ 44 คะแนน จากภาพที่ 4.28 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 44 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 25% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 7.5 ถึงต่ำกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 44 ถึง 54 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูงและนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 55 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 75% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยความยากข้อสอบ (ดูตารางที่ C2.3 ในภาคผนวก) ระดับขั้นสูงเท่ากับ 1.12 ลอจิทหรือคะแนนสเกลเท่ากับ 61 คะแนนและค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.17 ลอจิทหรือคะแนนสเกลเท่ากับ 48 คะแนนจากภาพที่ 4.28 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 9 คะแนนหรือมีคะแนนสเกลน้อยกว่า 48 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 22% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับ

นักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 9 ถึงต่ำกว่า 22 คะแนนหรือมีคะแนนสเกลตั้งแต่ 48 ถึง 60 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 22 คะแนนหรือมีคะแนนสเกลตั้งแต่ 61 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 78% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

จากภาพที่ 4.28 แสดงให้เห็นว่า ถ้าผู้วิจัยมีมุมมองต่อคะแนนจุดตัดขั้นต้นเป็นแบบมาตรฐานความสามารถและนำคะแนนจุดตัดทั้งสองค่าไปใช้ประเมินความสามารถนักเรียนจากกลุ่มตัวอย่างพบว่าในปีการศึกษา 2550 นักเรียนส่วนใหญ่จะมีความสามารถอยู่ในขั้นต่ำกว่าพื้นฐานและขั้นพื้นฐาน ในปีการศึกษา 2551 นักเรียนส่วนใหญ่อยู่ในขั้นพื้นฐาน และปีการศึกษา 2552 มีนักเรียนประมาณ 1% เท่านั้นที่อยู่ในขั้นสูง แต่ถ้าผู้วิจัยมีมุมมองแบบคะแนนผ่านขั้นต่ำสุดก็จะใช้คะแนนจุดตัดแรกเท่านั้นเพื่อแบ่งนักเรียนออกเป็นสองกลุ่มอย่างง่าย และพบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 37% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดของแต่ละปีการศึกษาได้ นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นร่วมของทั้งสามปี การศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 47 และ 57 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสำหรับการเรียนรู้ชีวิตกับสิ่งแวดล้อม

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ระดับขั้นพื้นฐานเท่ากับ .50 ลอจิท (เนื่องจากข้อสอบทุกข้อมีความยากอยู่ในระดับนี้) หรือคะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.29 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 2.5 คะแนนและนักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.2 ในภาคผนวก) ระดับขั้นสูงเท่ากับ .55 ลอจิทหรือคะแนนสเกลเท่ากับ 55 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.53 ลอจิทหรือคะแนนสเกลเท่ากับ 45 คะแนน จากภาพที่ 4.29 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 2.5 คะแนน (หมายถึงนักเรียนตอบข้อสอบไม่ถูกเลย) หรือมีคะแนนสเกลน้อยกว่า 45 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 25% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 2.5 ถึงต่ำกว่า 5 คะแนน (ตอบถูกสองถึงสามข้อ) หรือมีคะแนนสเกลตั้งแต่ 45 ถึง 54 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มี

ความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 5 คะแนน (ตอบถูกทุกข้อ) หรือมีคะแนนสเกลตั้งแต่ 55 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 75% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความ สามารถ
คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	
5	64	89.41%	7.5	62	17.25%	10.5	57	17.36%	ขั้นสูง
				55			56		
			5		67.94%			68.14%	ขั้นพื้นฐาน
2.5			2.5						
	50			45		2.5	47		ขั้นต่ำกว่า
		17.59%			14.81%			14.49%	
0	44		0	41		0	44		

ภาพที่ 4.29 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ชีวิตกับสิ่งแวดล้อม

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.3 ใน ภาคผนวก) ระดับขั้นสูงเท่ากับ .58 ลอจิทหรือคะแนนสเกลเท่ากับ 56 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.27 ลอจิทหรือคะแนนสเกลเท่ากับ 47 คะแนนจากภาพที่ 4.29 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 2.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 30% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 2.5 ถึงต่ำกว่า 10.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 47 ถึง 55 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบเท่ากับ 10.5 คะแนน (ตอบถูกทุกข้อ) หรือมีคะแนนสเกลตั้งแต่ 56 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 70% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

จากภาพที่ 4.29 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่าในปีการศึกษา 2550 นักเรียนส่วนใหญ่จะมีความสามารถสูงกว่าขั้นต่ำกว่าขั้นพื้นฐาน ในปีการศึกษา 2551 และ 2552 นักเรียนส่วนใหญ่อยู่ในขั้นพื้นฐาน นอกจากนี้ถ้า

ต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 16% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดของแต่ละปีการศึกษา นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นร่วมของทั้งสามปีการศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับ คะแนนสเกล 47 และ 56 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสำหรับการเรียนรู้สารและสมบัติของสาร

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ใน ภาคผนวก) ระดับขั้นสูงเท่ากับ .93 ลอจิทหรือคะแนนสเกลเท่ากับ 59 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ -.23 ลอจิทหรือคะแนนสเกลเท่ากับ 48 คะแนน จากภาพที่ 4.30 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 12.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 48 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 24% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 12.5 ถึงต่ำกว่า 20 คะแนนหรือมีคะแนนสเกลตั้งแต่ 48 ถึง 58 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 20 คะแนนหรือมีคะแนนสเกลตั้งแต่ 59 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 76% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ใน ภาคผนวก) ระดับขั้นพื้นฐานเท่ากับ .50 ลอจิท (เนื่องจากข้อสอบทุกข้อมีความยากอยู่ในระดับนี้) หรือคะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.30 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 10 คะแนนและนักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.3 ใน ภาคผนวก) ระดับขั้นสูงเท่ากับ 1.34 ลอจิทหรือคะแนนสเกลเท่ากับ 63 คะแนน และค่าเฉลี่ยความยากข้อสอบในระดับขั้นพื้นฐานเท่ากับ .11 ลอจิทหรือคะแนนสเกลเท่ากับ 51 คะแนน จากภาพที่ 4.30 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 10 คะแนนหรือมีคะแนนสเกลน้อยกว่า 51 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 23% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 10 ถึงต่ำกว่า 20 คะแนนหรือมีคะแนนสเกลตั้งแต่ 51 ถึง 62 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบเท่ากับ 20 คะแนน (ตอบถูกทุกข้อ) หรือมี

คะแนนสเกลเท่ากับ 63 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 77% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความสามารถ
คะแนนดิบ	คะแนนสเกล	ร้อยละ	คะแนนดิบ	คะแนนสเกล	ร้อยละ	คะแนนดิบ	คะแนนสเกล	ร้อยละ	
25	67	16.28%	20	67	75.15%			.61%	ขั้นสูง
20	59					20	63		ขั้นพื้นฐาน
		44.18%						42.15%	
12.5			10			10			ขั้นต่ำกว่า
	48			50			51		
		39.54%			24.85%			57.24%	
0	37		0	40		0	41		

ภาพที่ 4.30 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้สารและสมบัติของสาร

จากภาพที่ 4.30 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่าในปีการศึกษา 2550 นักเรียนส่วนใหญ่จะมีความสามารถขั้นพื้นฐาน ในปีการศึกษา 2551 มีจำนวนนักเรียน 3 ใน 4 ที่มีความสามารถสูงกว่าขั้นต่ำกว่าขั้นพื้นฐาน และในปีการศึกษา 2552 มีนักเรียนไม่ถึง 1% ที่มีความสามารถขั้นสูง นอกจากนี้ถ้าต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 41% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดของแต่ละปีการศึกษา นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นรวมของทั้งสามปีการศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 50 และ 61 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสาระการเรียนรู้แรงและการเคลื่อนที่

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขั้นพื้นฐานเท่ากับ .50 ลอจิท (เนื่องจากข้อสอบทุกข้อมีความยากอยู่ในระดับนี้) หรือ

คะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.31 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 5 คะแนนและนักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขั้นสูงเท่ากับ 1.03 ลอจิทหรือคะแนนสเกลเท่ากับ 60 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.13 ลอจิทหรือคะแนนสเกลเท่ากับ 49 คะแนน จากภาพที่ 4.31 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 2.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 49 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 24% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 2.5 ถึง 5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 49 ถึง 59 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 60 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 76% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.3 ในภาคผนวก) ขั้นสูงเท่ากับ 1.11 ลอจิทหรือคะแนนสเกลเท่ากับ 61 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ .01 ลอจิทหรือคะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.31 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 6.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 50 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 25% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 6.5 ถึงต่ำกว่า 14 คะแนนหรือมีคะแนนสเกลตั้งแต่ 50 ถึง 60 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบเท่ากับ 14 คะแนน (ตอบถูกทุกข้อ) หรือมีคะแนนสเกลเท่ากับ 61 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 75% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

จากภาพที่ 4.31 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่าในปีการศึกษา 2550 นักเรียนส่วนใหญ่จะมีความสามารถสูงกว่าขั้นต่ำ ในปีการศึกษา 2551 นักเรียนส่วนใหญ่จะมีความสามารถอยู่ในขั้นพื้นฐาน และในปีการศึกษา 2552 มีนักเรียนไม่ถึง 1.5% ที่มีความสามารถขั้นสูง นอกจากนี้ถ้าต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 33% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุดของแต่ละปีการศึกษา นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นร่วมของทั้งสามปี การศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 50 และ 61 คะแนนตามลำดับ

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความสามารถ
คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	คะแนนดิบ	คะแนน สเกล	ร้อยละ	
7.5	66	62.15%	7.5	65	18.93%			1.40%	ขั้นสูง
				60		14	61		
			5		67.47%			52.31%	ขั้นพื้นฐาน
5									
	50		2.5	49		6.5	50		
		37.85%			13.60%			46.29%	ขั้นต่ำกว่า
0	41		0	43		0	42		

ภาพที่ 4.31 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้แรงและการเคลื่อนที่

คะแนนจุดตัดขั้นต้นสาระการเรียนรู้พลังงาน

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขั้นสูงเท่ากับ .42 ลอจิตหรือคะแนนสเกลเท่ากับ 54 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.17 ลอจิตหรือคะแนนสเกลเท่ากับ 48 คะแนน จากภาพที่ 4.32 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 48 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 36% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 7.5 ถึงต่ำกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 48 ถึง 53 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบมากกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 54 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 64% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขั้นสูงเท่ากับ .52 ลอจิตหรือคะแนนสเกลเท่ากับ 55 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.82 ลอจิตหรือคะแนนสเกลเท่ากับ 42 คะแนน จากภาพที่ 4.32 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 2.5 คะแนนหรือมีคะแนนสเกล

น้อยกว่า 42 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 21% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 2.5 ถึง 10 คะแนนหรือมีคะแนนสเกลตั้งแต่ 42 ถึง 54 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบตั้งแต่ 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 55 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 79% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความสามารถ
คะแนนดิบ	คะแนน	ร้อยละ	คะแนนดิบ	คะแนน	ร้อยละ	คะแนนดิบ	คะแนน	ร้อยละ	
สเกล			สเกล			สเกล			
17.5	63	30.13%	15	61	19.43%	14	63	1.51%	ขั้นสูง
12.5			12.5						
	54			55			61		
		42.61			77.58%	12		76.41%	ขั้นพื้นฐาน
7.5	48		2.5	42		4	47		ขั้นต่ำกว่า
		30.13%			2.98%			22.08%	
0	39		0	39		0	43		

ภาพที่ 4.32 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้พลังงาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.3 ในภาคผนวก) ขั้นสูงเท่ากับ 1.09 ลอจิทหรือคะแนนสเกลเท่ากับ 61 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.34 ลอจิทหรือคะแนนสเกลเท่ากับ 47 คะแนน จากภาพที่ 4.32 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 4 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 19% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 4 ถึง 12 คะแนนหรือมีคะแนนสเกลตั้งแต่ 47 ถึง 60 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบเท่ากับ 14 คะแนน (ตอบถูกทุกข้อ) หรือมีคะแนน

สเกลเท่ากับ 61 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 81% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

จากภาพที่ 4.32 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่านักเรียนส่วนใหญ่จะมีความสามารถอยู่ในขั้นพื้นฐานทุกปีการศึกษา นอกจากนี้ถ้าต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 18% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุด นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นร่วมของทั้งสามปีการศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 46 และ 57 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขึ้นสูงเท่ากับ .50 ลอจิตหรือคะแนนสเกลเท่ากับ 55 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.25 ลอจิตหรือคะแนนสเกลเท่ากับ 47 คะแนน จากภาพที่ 4.33 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 32% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 7.5 ถึงต่ำกว่า 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 47 ถึง 54 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบตั้งแต่ 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 55 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 68% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขึ้นสูงเท่ากับ .48 ลอจิตหรือคะแนนสเกลเท่ากับ 55 คะแนน และค่าเฉลี่ยความยากข้อสอบขั้นพื้นฐานเท่ากับ -.26 ลอจิตหรือคะแนนสเกลเท่ากับ 47 คะแนน จากภาพที่ 4.33 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 47 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 32% ที่จะตอบถูกข้อสอบขั้นสูง สำหรับนักเรียนที่มีความสามารถขั้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 5 ถึง 10 คะแนนหรือมีคะแนนสเกลตั้งแต่ 42 ถึง 54 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขั้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นสูง และนักเรียนที่มีความสามารถขั้นสูงจะมีคะแนนดิบตั้งแต่ 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 55 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 68% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ชั้นพื้นฐานเท่ากับ .50 ลอจิต (เนื่องจากข้อสอบทุกข้อมีความยากอยู่ในระดับนี้) หรือ คะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.33 พบว่านักเรียนที่มีความสามารถชั้นต่ำกว่าชั้นพื้นฐานจะมีคะแนนต่ำกว่า 5 คะแนนและนักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบชั้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความสามารถ
คะแนนดิบ	คะแนนสเกล	ร้อยละ	คะแนนดิบ	คะแนนสเกล	ร้อยละ	คะแนนดิบ	คะแนนสเกล	ร้อยละ	
15	63	16.31%	15	62	14.29%	7.5	59	34.53%	ชั้นสูง
12.5			12.5						
	55			55					ชั้นพื้นฐาน
		43.00			65.19%	5			
7.5									ชั้นต่ำกว่า
	47		5	47			50		
		30.13%			20.52%	2.5		65.47%	
0	39		0	41		0	45		

ภาพที่ 4.33 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก

จากภาพที่ 4.33 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่า ในปีการศึกษา 2550 และ 2551 นักเรียนส่วนใหญ่จะมีความสามารถอยู่ในชั้นพื้นฐาน ส่วนปีการศึกษา 2552 นักเรียนส่วนใหญ่มีความสามารถในระดับสูงกว่าชั้นพื้นฐาน นอกจากนี้ถ้าต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 39% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุด นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นรวมของทั้งสามปีการศึกษาก็สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 48 และ 55 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสาระการเรียนรู้ดาราศาสตร์และอวกาศ

สำหรับปีการศึกษา 2550 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขึ้นสูงเท่ากับ .67 ลอจิตหรือคะแนนสเกลเท่ากับ 57 คะแนน และค่าเฉลี่ยความยากข้อสอบขึ้นพื้นฐานเท่ากับ -.45 ลอจิตหรือคะแนนสเกลเท่ากับ 45 คะแนน จากภาพที่ 4.34 พบว่านักเรียนที่มีความสามารถขึ้นต่ำกว่าขึ้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 45 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 25% ที่จะตอบถูกข้อสอบขึ้นสูง สำหรับนักเรียนที่มีความสามารถขึ้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 5 ถึงต่ำกว่า 10 คะแนนหรือมีคะแนนสเกลตั้งแต่ 45 ถึง 56 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขึ้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขึ้นสูง และนักเรียนที่มีความสามารถขึ้นสูงจะมีคะแนนดิบตั้งแต่ 10 คะแนนหรือมีคะแนนสเกลตั้งแต่ 57 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 75% ที่จะตอบถูกข้อสอบขึ้นพื้นฐาน

สำหรับปีการศึกษา 2551 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขึ้นสูงเท่ากับ .38 ลอจิตหรือคะแนนสเกลเท่ากับ 54 คะแนน และค่าเฉลี่ยความยากข้อสอบขึ้นพื้นฐานเท่ากับ -.42 ลอจิตหรือคะแนนสเกลเท่ากับ 46 คะแนน จากภาพที่ 4.34 พบว่านักเรียนที่มีความสามารถขึ้นต่ำกว่าขึ้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7.5 คะแนนหรือมีคะแนนสเกลน้อยกว่า 46 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 31% ที่จะตอบถูกข้อสอบขึ้นสูง สำหรับนักเรียนที่มีความสามารถขึ้นพื้นฐานจะมีคะแนนดิบตั้งแต่ 7.5 ถึง 10 คะแนนหรือมีคะแนนสเกลตั้งแต่ 46 ถึง 53 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 50% ที่จะตอบถูกข้อสอบขึ้นพื้นฐานและมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขึ้นสูง และนักเรียนที่มีความสามารถขึ้นสูงจะมีคะแนนดิบตั้งแต่ 12.5 คะแนนหรือมีคะแนนสเกลตั้งแต่ 54 คะแนนขึ้นไป นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 69% ที่จะตอบถูกข้อสอบขึ้นพื้นฐาน

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.3 ในภาคผนวก) ขึ้นสูงและขึ้นพื้นฐานถูกรวมเป็นชั้นเดียวกัน เนื่องจากข้อสอบข้อที่ 36 เป็นข้อสอบยากและเป็นข้อเดียวที่อยู่ในระดับขึ้นสูง ทำให้ค่าความยากเฉลี่ยข้อสอบมีค่ามากกว่าค่าสูงสุดของค่าความสามารถนักเรียน ค่าเฉลี่ยความยากข้อสอบขึ้นพื้นฐานเท่ากับ -.34 ลอจิตหรือคะแนนสเกลเท่ากับ 53 คะแนน จากภาพที่ 4.34 พบว่านักเรียนที่มีความสามารถขึ้นต่ำกว่าขึ้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 7 คะแนนหรือมีคะแนนสเกลน้อยกว่า 53 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 33% ที่จะตอบถูกข้อสอบขึ้นสูง สำหรับนักเรียนที่มีความสามารถขึ้นพื้นฐานจะมีคะแนนดิบมากกว่า 7 คะแนนขึ้นไปหรือมีคะแนนสเกลตั้งแต่ 53 คะแนน นักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างน้อยที่สุด 67% ที่จะตอบถูกข้อสอบขึ้นพื้นฐาน

ปีการศึกษา 2550			ปีการศึกษา 2551			ปีการศึกษา 2552			ระดับความสามารถ
คะแนนดิบ	คะแนน	ร้อยละ	คะแนนดิบ	คะแนน	ร้อยละ	คะแนนดิบ	คะแนน	ร้อยละ	

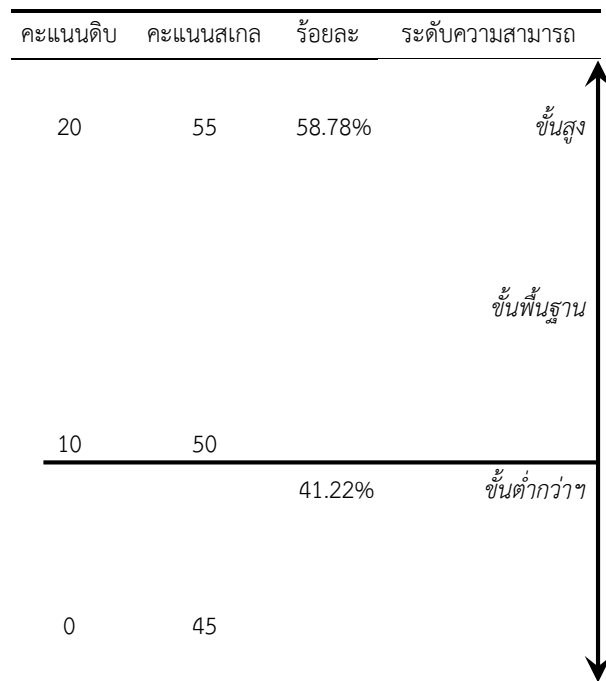
สเกล			สเกล			สเกล		
12.5	64	24.92%	17.5	65	33.52%	0%	↑ ขั้นสูง	
10			12.5					
57			54					
		49.64			41.09%	14	58	41.28% ↑ ขั้นพื้นฐาน
5			7.5					
45			46			7	53	
		25.44%			25.39%			58.72% ↓ ขั้นต่ำกว่า
0	38		0	38		0	45	

ภาพที่ 4.34 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ดาราศาสตร์และอวกาศ

จากภาพที่ 4.34 แสดงให้เห็นว่าสามารถนำคะแนนจุดตัดขั้นต้นทั้งสองระดับมากำหนดเป็นมาตรฐานความสามารถ พบว่าในปีการศึกษา 2550 และ 2551 นักเรียนส่วนใหญ่มีความสามารถขั้นพื้นฐาน แต่ปีการศึกษา 2552 นักเรียนส่วนใหญ่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐาน นอกจากนี้ถ้าต้องการกำหนดคะแนนผ่านขั้นต่ำสุดก็สามารถใช้คะแนนจุดตัดแรกเพื่อแบ่งนักเรียนออกเป็นสองกลุ่ม พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 37% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุด นอกจากนี้เนื่องจากคะแนนจุดตัดของแต่ละระดับค่อนข้างคงเส้นคงวา ดังนั้นถ้าผู้วิจัยต้องการกำหนดคะแนนจุดตัดขั้นต้นร่วมของทั้งสามปีการศึกษาที่สามารถคำนวณจากค่าเฉลี่ยคะแนนจุดตัดซึ่งเท่ากับคะแนนสเกล 48 และ 56 คะแนนตามลำดับ

คะแนนจุดตัดขั้นต้นสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

สำหรับปีการศึกษา 2552 ค่าเฉลี่ยค่าประมาณความยากข้อสอบ (ดูตารางที่ C2.1 ในภาคผนวก) ขั้นพื้นฐานเท่ากับ .50 ลอจิต (เนื่องจากข้อสอบทุกข้อมีความยากอยู่ในระดับนี้) หรือคะแนนสเกลเท่ากับ 50 คะแนน จากภาพที่ 4.35 พบว่านักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานจะมีคะแนนดิบต่ำกว่า 10 คะแนนและนักเรียนในระดับนี้จะมีโอกาสเฉลี่ยอย่างมากที่สุด 49% ที่จะตอบถูกข้อสอบขั้นพื้นฐาน เมื่อกำหนดคะแนนผ่านขั้นต่ำสุด พบว่าจะมีจำนวนนักเรียนเฉลี่ยประมาณ 41% ที่ไม่ผ่านเกณฑ์ขั้นต่ำสุด



ภาพที่ 4.35 คะแนนจุดตัดขั้นต้นความสามารถสาระการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

บทที่ 5

สรุปผล อภิปรายผลและข้อเสนอแนะ

โครงการวิจัยนี้ นำผลการทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Educational Testing: ONET) ไปประยุกต์ใช้ในสถานการณ์ที่แตกต่างจากที่คุ้นเคย ผลลัพธ์เบื้องต้นของโครงการวิจัยนี้ นอกจากจะได้คะแนนจุดตัดขั้นต้นแล้ว ยังจะได้แนวคิดและวิธีการใหม่สำหรับพัฒนาโครงสร้างของตัวแปรความสามารถทางวิทยาศาสตร์ด้วยวิธีทำแผนที่ภาวะสันนิษฐาน และใช้แผนที่ดังกล่าวในกระบวนการกำหนดคะแนนจุดตัด ผลลัพธ์ระยะไกลของโครงการวิจัยนี้ มุ่งหวังเป็นอย่างยิ่งว่าคะแนนจุดตัดขั้นต้นและแผนที่ภาวะสันนิษฐานจะให้สารสนเทศที่เป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดจริงบนพื้นฐานของข้อมูลที่ต้องการและเพียงพอ และเมื่อกำหนดคะแนนจุดตัดจริงแล้ว คะแนนจุดตัดจริงจะเป็นประโยชน์ต่อการสร้างมาตรฐานความสามารถทางวิทยาศาสตร์ของนักเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 รวมทั้งสถาบันทดสอบทางการศึกษาแห่งชาติจะนำแผนที่ภาวะสันนิษฐานไปใช้ปรับปรุงกระบวนการสร้างและพัฒนาแบบสอบ ONET ให้ดียิ่งขึ้น โดยขยายแนวคิดไปสู่วิชาอื่นและในระดับชั้นอื่นด้วย

ผลการวิจัยที่เกิดขึ้นมีสาระสำคัญหลายส่วน แต่ละส่วนบรรยายจุดเด่นและข้อจำกัดที่ค้นพบจากโครงการวิจัยนี้ ดังนั้นเนื้อหาในบทนี้ นอกจากจะสรุปผลการวิจัยแล้ว ยังอภิปรายผลการวิจัยในประเด็นต่างๆ ตามประโยชน์ที่ได้รับจากการวิจัยในครั้งนี้ และให้ข้อเสนอแนะเพื่อนำผลวิจัยไปใช้ได้ อย่างถูกต้องตามขอบเขตของโครงการวิจัย รวมทั้งข้อเสนอแนะเพื่อการค้นคว้าวิจัยในอนาคต เนื้อหาแต่ละส่วนมีรายละเอียดดังต่อไปนี้

สรุปผลวิจัย

1) แผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์ทั้งแบบภาพรวมและรายละเอียดการเรียนรู้ที่พัฒนาขึ้นจากข้อสอบ ONET วิชาวิทยาศาสตร์ชั้นประถมศึกษาปีที่ 6 ตั้งแต่ปีการศึกษา 2550 ถึง 2552 จำนวน 117 ข้อ ประกอบด้วยโครงสร้างความสามารถนักเรียนห้าระดับ ตั้งแต่ขั้นต่ำกว่าขั้นต้น (Below Beginning) ซึ่งเป็นขั้นต่ำสุด ไปจนถึงขั้นต้น (Beginning) ขั้นพื้นฐาน (Basic) ขั้นเชี่ยวชาญ (Proficient) และขั้นสูง (Advanced) ซึ่งเป็นขั้นสูงสุดของโครงสร้างนี้ แต่ละระดับมีคำบรรยายเฉพาะที่เกิดจากการวิเคราะห์ข้อมูลคุณสมบัติข้อสอบ (item properties) และข้อมูลเกี่ยวกับมาตรฐานการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 คำบรรยายเหล่านี้จะสะท้อนความสามารถนักเรียนที่แตกต่างกันในแต่ละระดับตามแผนที่ภาวะสันนิษฐานทั้งเนื้อหาวิทยาศาสตร์ที่นักเรียนต้องรู้และทักษะการคิดที่ใช้แก้ปัญหาที่ซับซ้อนแตกต่างกัน

ผลการตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานเชิงทฤษฎีด้วยข้อมูลคำตอบข้อสอบ ONET โดยใช้กรอบโมเดลคะแนนคำตอบหลายค่าที่มีสัมประสิทธิ์แบบสุ่มจากการวัดหลายมิติ (Multidimensional Random Coefficients Multinomial Logit (MRCML) model: Adams, Wilson, & Wu, 1997) พบว่า แผนที่ภาวะสันนิษฐานเชิงทฤษฎีสอดคล้องกับข้อมูลคำตอบแตกต่างกันและนำไปสู่การรวมระดับความสามารถเหลือสามระดับเพื่อให้โครงสร้างความสามารถใหม่บรรยายความสามารถนักเรียนได้เหมาะสมยิ่งขึ้นและเรียกโครงสร้างใหม่นี้ว่า แผนที่ภาวะสันนิษฐานเชิงประจักษ์ความสามารถทางวิทยาศาสตร์ ประกอบด้วยความสามารถนักเรียนขั้นต่ำกว่าขั้นพื้นฐาน (หรือขั้นต่ำกว่าขั้นต้นเดิม) ขั้นพื้นฐาน (รวมขั้นต้นและขั้นพื้นฐาน) และขั้นสูง (รวมขั้นเชี่ยวชาญและขั้นสูง)

2) โมเดลการวัดแบบเอกมิติ (unidimensional model) และพหุมิติ (multidimensional model) ของแต่ละปีการศึกษาถูกพัฒนาขึ้นตามโครงสร้างความสามารถทางวิทยาศาสตร์ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ข้อกำหนดลักษณะแบบสอบ ONET (test specification) และจากผลวิเคราะห์คุณสมบัติข้อสอบ ONET โดยผู้เชี่ยวชาญ เพื่อใช้ประมาณค่าความสามารถนักเรียนแต่ละคนและค่าความยากข้อสอบแต่ละข้อสำหรับใช้กำหนดคะแนนจุดตัดขั้นต้น

ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้งหมดกับข้อมูลคำตอบข้อสอบ ONET ของแต่ละปีการศึกษาด้วยกรอบโมเดล MRCML พบว่าโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดในปีการศึกษา 2550 คือโมเดลพหุมิติระหว่างข้อสอบ BimFit-50 ปีการศึกษา 2551 คือโมเดลเอกมิติ UniFit-51 และปีการศึกษา 2552 คือโมเดลพหุมิติภายในข้อสอบที่ให้คะแนนหลายค่า WimFit-52P

เนื่องจากโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดเหล่านี้นี้อธิบายคำตอบคำถามวิจัยเพียงบางส่วน ผู้วิจัยจึงคัดเลือกโมเดลเพิ่มเติมอีกสามโมเดลเพื่อให้ผลวิเคราะห์ตอบคำถามวิจัยได้ครบถ้วน ได้แก่ โมเดลเอกมิติ UniFit-50 สำหรับปีการศึกษา 2550 โมเดลพหุมิติระหว่างข้อสอบ BimFit-51 สำหรับปีการศึกษา 2551 และโมเดลเอกมิติที่ให้คะแนนหลายค่า UniFit-51P สำหรับปีการศึกษา 2552 ก่อนที่โมเดลทั้งหมดจะถูกนำไปใช้ประมาณค่าพารามิเตอร์ต่างๆ ผู้วิจัยได้ตรวจสอบความตรงข้ามกลุ่มของโมเดลพบว่า ค่าพารามิเตอร์ที่ประมาณได้จากโมเดลโดยใช้กลุ่มตัวอย่างอีกชุดหนึ่งคงเส้นคงวา มีผลต่างของค่าพารามิเตอร์เชิงโครงสร้างระหว่างกลุ่มตัวอย่างสอดคล้องกลมกลืน (fitting sample) และกลุ่มตัวอย่างความตรงข้ามกลุ่ม (validating sample) น้อยมาก

3) เมื่อรวมกลุ่มตัวอย่างสอดคล้องกลมกลืนและกลุ่มตัวอย่างความตรงข้ามกลุ่มเข้าด้วยกัน และนำไปใช้ประมาณค่าความสามารถนักเรียนและค่ายากข้อสอบในแต่ละปีการศึกษาตามโมเดลการวัดทั้งหมดโมเดลข้างต้น จะได้สเกลคะแนนความสามารถนักเรียน เมื่อนำสเกลดังกล่าวไปเชื่อมโยงกับแผนที่ภาวะสันนิษฐานเชิงประจักษ์และค่าเฉลี่ยค่าความยากข้อสอบในแต่ละระดับความยากตามแผนที่ภาวะสันนิษฐานเชิงประจักษ์ ก็จะได้คะแนนจุดตัดขั้นต้นทั้งในภาพรวมและรายสาระการเรียนรู้

ผลการกำหนดคะแนนจุดตัดขั้นต้นในภาพรวมพบว่า หลังจากที่ได้แปลงค่าประมาณความสามารถเป็นคะแนนสเกลที่มีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 แล้ว

คะแนนสเกลที่ใช้แบ่งระดับความสามารถนักเรียนระหว่างขั้นต่ำกว่าขั้นพื้นฐานกับขั้นพื้นฐานคือ 47, 46, และ 47 คะแนน และคะแนนสเกลที่ใช้แบ่งระดับความสามารถนักเรียนระหว่างขั้นพื้นฐานและขั้นสูงคือ 59, 56, และ 57 ตามลำดับตั้งแต่ปีการศึกษา 2550 ถึง 2552 เมื่อวิเคราะห์ผลกระทบที่เกิดขึ้นเมื่อนำคะแนนจุดตัดไปใช้ประเมินความสามารถนักเรียนพบว่า ในภาพรวมทั้งสามปีการศึกษา จะมีนักเรียนร้อยละ 35 ที่มีความสามารถอยู่ในขั้นต่ำกว่าขั้นพื้นฐาน ร้อยละ 49 อยู่ในขั้นพื้นฐาน และร้อยละ 16 อยู่ในขั้นสูง

ผลการกำหนดคะแนนจุดตัดขั้นต้นรายละเอียดการเรียนรู้พบว่า คะแนนสเกลที่ใช้แบ่งระดับความสามารถนักเรียนระหว่างขั้นต่ำกว่าขั้นพื้นฐานกับขั้นพื้นฐานจากทุกสาระการเรียนรู้มีค่าตั้งแต่ 46 ถึง 50 คะแนน และคะแนนสเกลที่ใช้แบ่งระดับความสามารถนักเรียนระหว่างขั้นพื้นฐานและขั้นสูงจากทุกสาระการเรียนรู้มีค่าตั้งแต่ 55 ถึง 61 คะแนน และเมื่อนำคะแนนจุดตัดขั้นต้นรายละเอียดการเรียนรู้เหล่านี้ไปใช้จะเกิดผลกระทบคือ โดยภาพรวมจะมีนักเรียนร้อยละ 33 ที่มีความสามารถอยู่ในขั้นต่ำกว่าขั้นพื้นฐานของแต่ละสาระการเรียนรู้

อภิปรายผลวิจัย

1) ข้อค้นพบที่สำคัญของโครงการวิจัยนี้คือวิธีการเชื่อมต่อกะบวนพัฒนาแบบสอบ ONET เข้ากับกระบวนการกำหนดคะแนนจุดตัดด้วยแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์ โดยผู้วิจัยประยุกต์จากกรอบการสร้างโมเดลภาวะสันนิษฐาน (Construct Modeling framework: Wilson, 2005) ปัญหาทั่วไปที่เกิดขึ้นในกระบวนการกำหนดคะแนนจุดตัดคือ นอกจากขั้นตอนตัดสินว่าจะใช้ค่าใดเป็นคะแนนจุดตัดจริงนั้นจะขึ้นอยู่กับกรณีการวินิจฉัยและข้อตกลงร่วมกับของคณะบุคคลซึ่งอาจจะไม่เป็นปรนัย (subjective) แล้ว ข้อมูลเกี่ยวกับคะแนนสอบที่มอบแก่ผู้มีอำนาจเหล่านั้นพิจารณาอาจจะไม่เพียงพอและขึ้นอยู่กับค่าสถิติพื้นฐานของคะแนนสอบเป็นหลัก โดยที่ผู้มีอำนาจตัดสินไม่สามารถตีความคะแนนสอบเหล่านั้นว่าสัมพันธ์กับความสามารถนักเรียนอย่างไร

นิยามของการวัดของ Stevens (1946) ระบุว่า การวัดคือกระบวนการกำหนดค่าเชิงปริมาณให้กับคุณลักษณะที่สนใจ นิยามดังกล่าวนี้แสดงให้เห็นว่าคะแนนสอบ ONET ที่แตกต่างกันจะสะท้อนคุณลักษณะของนักเรียนไม่เหมือนกัน และค่าเชิงปริมาณก็ไม่ได้เกิดขึ้นในขั้นตอนรวมคะแนนรายข้อเป็นคะแนนรวมแล้วจึงคำนวณค่าสถิติพื้นฐาน ในความเป็นจริงแล้วค่าเชิงปริมาณเกิดขึ้นตั้งแต่ขั้นตอนให้คะแนนคำตอบข้อสอบ ONET รายข้อ (ก่อนที่จะรวมเป็นคะแนนและคำนวณค่าสถิติพื้นฐาน) ธรรมชาติของการกำหนดค่าเชิงปริมาณให้กับคำตอบของนักเรียน (ซึ่งคำตอบจะสัมพันธ์กับความสามารถของนักเรียน นักเรียนที่รอบรู้จะมีรูปแบบคำตอบแบบหนึ่ง ในขณะที่นักเรียนที่ไม่รอบรู้จะมีรูปแบบคำตอบอีกแบบหนึ่ง) เป็นแนวคิดแบบอิงเกณฑ์ (criterion-referenced) คือข้อสอบ ONET แต่ละข้อถูกสร้างขึ้นเพื่อวัดเนื้อหาและทักษะการคิดเฉพาะ คำตอบของนักเรียนจึงสะท้อนให้เห็นว่านักเรียนบรรลุในสิ่งที่ข้อสอบ ONET ขอนั้นวัดหรือไม่ ถ้าบรรลุก็จะได้คะแนน หรือไม่บรรลุก็จะ

ไม่ได้คะแนน การให้ค่าเชิงปริมาณกับคำตอบข้อสอบ ONET แต่ละข้อจึงเป็นแนวคิดแบบอิงเกณฑ์โดยใช้ความถูกต้องของคำตอบจำแนกนักเรียน ความเข้าใจคลาดเคลื่อนเกิดขึ้นเมื่อนำคะแนนแต่ละข้อรวมเป็นคะแนนรวมแล้วใช้แนวคิดแบบอิงกลุ่ม (norm-referenced) เพื่อตีความหมายของคะแนนรวม ความไม่สอดคล้องกันของแนวคิดทั้งสองในกระบวนการพัฒนาแบบสอบและวิเคราะห์คะแนนสอบส่งผลให้เมื่อนำคะแนนรวมไปใช้กำหนดคะแนนจุดตัดจะไม่สามารถตีความหมายของคะแนนเป็นความสามารถนักเรียนตามเป้าหมายที่แบบสอบหรือข้อสอบต้องการวัด

ความสามารถนักเรียนไม่ควรตีความแบบอิงกลุ่ม (norm-referenced approach) เช่นนักเรียนที่ได้คะแนนสอบ ONET วิชาวิทยาศาสตร์ต่ำกว่าค่าเฉลี่ยอยู่สองเท่าของค่าคลาดเคลื่อนมาตรฐาน เป็นต้น เพราะค่าสถิติพื้นฐานเหล่านี้ไม่ได้ให้ข้อมูลที่แม่นยำมากพอและอาจจะไม่สัมพันธ์กับคะแนนรายข้อของนักเรียน รวมทั้งไม่สัมพันธ์กับโครงสร้างระดับความสามารถ (เพราะในขั้นตอนให้คะแนนคำตอบรายข้อ คะแนนที่นักเรียนได้รับก็ไม่ได้กำหนดว่าถ้าคำตอบของนักเรียนส่วนน้อยแตกต่างจากนักเรียนส่วนใหญ่จะได้คะแนนหรือไม่ได้คะแนน) และเมื่อผู้มีอำนาจตัดสินใจได้รับผลวิเคราะห์เหล่านี้ไปใช้กำหนดคะแนนจุดตัดก็จะไม่สามารถตีความคะแนนเหล่านั้นได้ว่านักเรียนที่มีคะแนนต่ำกว่าค่าเฉลี่ยอยู่สองเท่าของค่าความคลาดเคลื่อนมาตรฐานหมายถึงนักเรียนมีความรู้อะไรและมีทักษะการคิดอะไรบ้าง

ผลการวิจัยสะท้อนให้เห็นว่าเมื่อวิเคราะห์และตรวจสอบโครงสร้างความสามารถที่ข้อสอบ ONET ต้องการวัดและนำโครงสร้างนั้นที่อยู่ในรูปของแผนที่ภาวะสันนิษฐานไปใช้ในกระบวนการให้คะแนนคำตอบ การประมาณค่าความสามารถนักเรียน และการตีความคะแนนสอบที่เชื่อมโยงกับแผนที่ภาวะสันนิษฐานในตอนต้น จะทำให้คะแนนที่เสนอให้เป็นคะแนนจุดตัดขั้นต้นก่อนเข้าสู่กระบวนการกำหนดคะแนนจุดตัดจริงสามารถตีความให้สัมพันธ์กับความสามารถของนักเรียนที่ใช้ทำข้อสอบ ONET ดังนั้นคะแนนสอบ ONET ไม่ว่าจะเป็นคะแนนรวมทั้งฉบับหรือรายละเอียดการเรียนรู้สามารถตีความได้ว่านักเรียนที่มีคะแนนอยู่ในช่วงต่างๆ มีแนวโน้มที่มีความสามารถอยู่ในขั้นต่ำกว่าขั้นพื้นฐาน ขั้นพื้นฐานหรือขั้นสูง เมื่อนำสารสนเทศนี้เข้าไปใช้ในกระบวนการกำหนดคะแนนจุดตัดจริงก็จะช่วยให้ผู้มีอำนาจตัดสินใจเข้าใจความหมายของคะแนน ONET ทั้งแบบอิงเกณฑ์ (ช่วงคะแนนที่สัมพันธ์กับโครงสร้างของสิ่งที่ข้อสอบมุ่งวัด) และแบบอิงกลุ่ม (ค่าสถิติพื้นฐาน เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และเปอร์เซ็นต์ไทล์) ก่อนที่จะอภิปรายลงความเห็นเพื่อสรุปว่าควรใช้ค่าใดเป็นคะแนนจุดตัดจริง

2) การพัฒนาแผนที่ภาวะสันนิษฐานความสามารถทางวิทยาศาสตร์จากข้อสอบ ONET เป็นการใช้กรอบการสร้างโมเดลภาวะสันนิษฐานเพื่อแก้ไขปัญหาสถานการณ์ทดสอบแบบดั้งเดิมที่ละเลยการพัฒนาโครงสร้างความสามารถทางวิทยาศาสตร์ เช่นเดียวกับโมเดลสามเหลี่ยมการประเมิน (the Assessment triangle: NRC, 2001) และการออกแบบการประเมินที่เน้นหลักฐาน (Evidence-centered design: Mislevy, Steinberg, & Almond, 2003) ที่มุ่งตอบคำถามว่า “สิ่งที่แบบสอบ ONET ต้องการวัดจริงๆ นั้นคืออะไร” คำตอบที่เหมาะสมของคำถามข้างต้นนำไปสู่การ

ออกแบบการทดสอบที่จะต้องให้สารสนเทศเกี่ยวกับนักเรียนมากขึ้น เพราะการสอบ ONET ต้องการข้อมูลที่บ่งชี้ได้ว่านักเรียนมีความรู้มากน้อยเพียงใด สามารถทำอะไรได้บ้าง นักเรียนที่มีผลสัมฤทธิ์บรรลุตามมาตรฐานการเรียนรู้เป็นอย่างไร ในขณะที่เดียวกันนักเรียนที่ไม่สามารถบรรลุได้นั้นมีลักษณะอย่างไร สารสนเทศเหล่านี้ไม่ได้มุ่งไปสู่การสรุปรวมนักเรียนมีคะแนนมากกว่าค่าเฉลี่ยหรือมีคะแนนอยู่ที่เปอร์เซ็นต์ไทล์ที่ 80 ขึ้นไป (เพราะไม่มีทราบว่ามีคะแนนเท่ากับค่าเฉลี่ย หรืออยู่ที่ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 80 มีความสามารถอย่างไร และนักเรียนทั้งสองแบบต่างกันอย่างไร) แต่มุ่งหวังให้นักเรียนที่ไม่บรรลุตามมาตรฐานการเรียนรู้สามารถพัฒนาตนเองให้ดียิ่งขึ้น ครูจะรู้ว่านักเรียนที่มีคะแนนแตกต่างกันหมายถึงมีความสามารถแตกต่างกันอย่างไร ดังนั้นคำตอบที่ต้องการเหล่านี้จะต้องมาจากการพิจารณาไตร่ตรองอย่างละเอียดถี่ถ้วนตั้งแต่เริ่มต้นกระบวนการทดสอบว่า “สิ่งที่การทดสอบ ONET ต้องการวัดจริงๆ นั้นคืออะไร”

ข้อดีของการออกแบบโครงสร้างของสิ่งที่ต้องการวัดให้ชัดเจนมากที่สุดเท่าที่จะเป็นไปได้คือ นอกจากจะให้สารสนเทศตามที่ได้กล่าวไว้ข้างต้นแล้ว สารสนเทศที่เกิดขึ้นในกระบวนการทดสอบยังสามารถใช้เป็นหลักฐานสนับสนุนความตรงของผลการทดสอบได้ทั้งหมดตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยาของประเทศสหรัฐอเมริกา (the Standards for Educational and Psychological Testing: AERA, APA, & NCME, 2014) แต่สำหรับการทดสอบ ONET ซึ่งไม่ได้กำหนดหรือออกแบบโครงสร้างของสิ่งที่วัดให้ชัดเจนก่อนนั้น เมื่อนำกรอบการสร้างโมเดลภาวะสันนิษฐานมาประยุกต์ใช้ทำให้เกิดข้อดีที่สำคัญคือ รู้วิธีการสร้างและพัฒนาโครงสร้างของสิ่งที่ต้องการวัดจากข้อสอบที่มีอยู่แล้ว แล้วใช้โครงสร้างที่พัฒนาขึ้นสำหรับสร้างข้อสอบในอนาคตต่อไป แต่อย่างไรก็ตามวิธีดังกล่าวก็มีข้อจำกัดด้วยคือ จำเป็นต้องใช้ข้อสอบจำนวนมากเพื่อวิเคราะห์สาระสำคัญที่มีร่วมกันจากข้อสอบเหล่านั้น ข้อสอบจำนวนน้อยจะไม่สามารถให้ข้อมูลเพียงพอต่อการสร้างโครงสร้างของสิ่งที่สนใจจะวัด รวมทั้งโครงสร้างของสิ่งที่สนใจจะวัดที่พัฒนาขึ้นไม่ว่าจะมาจากข้อสอบจำนวนน้อยหรือจำนวนมากจะไม่สามารถอธิบายรายละเอียดทั้งหมดของข้อสอบได้ทุกข้อ

3) ตรรกะการพัฒนาคะแนนจุดตัดไม่ได้อยู่วิธีกำหนดคะแนนจุดตัดเพียงอย่างเดียว Zieky, Perie และ Livingston (2008) กล่าวอย่างชัดเจนว่า นักวิจัยจะต้องเลือกวิธีกำหนดคะแนนจุดตัดที่เหมาะสมกับบริบท เงื่อนไขและวัตถุประสงค์ของการกำหนดคะแนนจุดตัดนั้นด้วย รวมทั้ง Perie (2008) เน้นย้ำว่าปัญหาความไม่สมบูรณ์และความไม่ถูกต้องของคะแนนจุดตัด ส่วนหนึ่งเกิดขึ้นจากการไม่ให้ความสำคัญมากพอในการพัฒนาคำบรรยายระดับความสามารถ (performance level descriptor) ก่อนที่จะเริ่มกระบวนการกำหนดคะแนนจุดตัด และไม่ใช้คำบรรยายระดับความสามารถตั้งแต่ต้นของกระบวนการสร้างและพัฒนาข้อสอบทำให้คะแนนจุดตัดที่เกิดขึ้นขาดหลักฐานสนับสนุน คะแนนจุดตัดและไม่สามารถอธิบายความสัมพันธ์ระหว่างคะแนนจุดตัดกับระดับความสามารถที่มุ่งหวังและความหมายของคะแนนที่ใช้กำหนดคะแนนจุดตัด เมื่อเปรียบเทียบผลลัพธ์จากกระบวนการกำหนดคะแนนจุดตัดที่เกิดขึ้นในปัจจุบันกับตรรกะการพัฒนาคะแนนจุดตัดของ Wilson (2015) พบว่าให้สารสนเทศเพียงบางส่วนเท่านั้น โดยเฉพาะคะแนนจุดตัดขาดการเชื่อมโยงกับ “สิ่งที่

แบบสอบนั้นๆ ต้องการวัดจริงๆ” ขาดหลักฐานสนับสนุนความตรงของคะแนนจุดตัด ไม่สามารถแปลความหมายคะแนนจุดตัดได้ และนำคะแนนสอบไปใช้กำหนดคะแนนจุดตัดไม่เหมาะสม

ข้อค้นพบของโครงการวิจัยนี้ไม่ได้ทำให้หลักการกำหนดคะแนนจุดตัดเปลี่ยนแปลง แต่เป็นการเสนอแนวทางเพื่อสนับสนุนวิธีการกำหนดคะแนนจุดตัดที่มีอยู่แล้วเกือบ 100 วิธีให้สามารถกำหนดค่าเพื่อที่จะเป็นจุดตัดได้ถูกต้องและน่าเชื่อถือมากยิ่งขึ้น โดยใช้แผนที่ภาวะสันนิษฐานแทนคำบรรยายระดับความสามารถนักเรียนในการออกแบบกระบวนการประเมิน สนับสนุนกระบวนการกำหนดคะแนนจุดตัดและเชื่อมโยงข้อมูลที่เกี่ยวข้องเข้าด้วยกัน นั่นหมายความว่าหลังจากข้อมูลเกี่ยวข้องกับคะแนนสอบและข้อมูลเกี่ยวกับผู้สอบถูกเชื่อมโยงกันหมดด้วยแผนที่ภาวะสันนิษฐาน ผู้มีอำนาจตัดสินใจสามารถเลือกวิธีการกำหนดคะแนนจุดตัดที่คิดว่าเหมาะสมกับบริบทของการทดสอบ ONET อื่นได้ เช่น อาจจะใช้วิธีของ Angoff หรือวิธี Bookmark แต่ข้อมูลคะแนนสอบ ONET ที่ผู้มีอำนาจตัดสินใจจะได้รับนั้นจะอยู่ในรูปของแผนที่ภาวะสันนิษฐาน ทำให้ผู้พิจารณาเห็นความเชื่อมโยงกันของข้อมูลทั้งหมดและความหมายของคะแนน ONET ด้วย

4) แผนที่ภาวะสันนิษฐานเชิงประจักษ์ของโครงการวิจัยนี้ที่ประกอบด้วยความสามารถนักเรียนสามารถบ่งชี้ไม่ได้ว่าเป็นโครงสร้างที่ดีที่สุด แต่เป็นโครงสร้างที่เหมาะสมที่สุด โครงสร้างหนึ่งที่เกิดขึ้นจากการวิเคราะห์ข้อมูลข้อสอบ ONET วิชาวิทยาศาสตร์ปีการศึกษา 2550 ถึง 2552 เป็นหลัก วิธีการที่ผู้วิจัยใช้อยู่บนพื้นฐานของหลักฐานเชิงประจักษ์ ผู้วิจัยจึงมั่นใจว่าแผนที่ภาวะสันนิษฐานเชิงประจักษ์ที่พัฒนาขึ้นนี้สะท้อนสถานการณ์การทดสอบในปัจจุบัน เมื่อผู้ออกข้อสอบ ONET นำแผนที่ภาวะสันนิษฐานไปใช้เป็นโครงสร้างตั้งต้น ข้อสอบ ONET ที่จะสร้างขึ้นใหม่จะมีหลักฐานมาสนับสนุนความตรงเชิงโครงสร้างของสิ่งที่ต้องการวัดมากขึ้น (เพราะโครงสร้างมาจากข้อสอบ ONET ในอดีต) ตัวอย่าง ถ้าต้องการสร้างข้อสอบวัดความสามารถขั้นสูงตามแผนที่ภาวะสันนิษฐานเชิงประจักษ์ ผู้สร้างข้อสอบ ONET จะรู้โดยทันทีว่านักเรียนที่มีความสามารถขั้นสูงนั้นมีคุณลักษณะอย่างไร ข้อสอบที่ผ่านมาสำหรับวัดความสามารถของนักเรียนในระดับนี้เป็นอย่างไร มีลีลาการตั้งโจทย์คำถามแบบใด ใช้ข้อมูลประเภทใดประกอบคำถาม สามารถบรรยายแนวโน้มได้ว่านักเรียนที่มีความสามารถขั้นสูงจะตอบแบบใด และนักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นสูงจะตอบแบบใดด้วย แผนที่ภาวะสันนิษฐานจึงช่วยชี้แนะผู้ออกข้อสอบว่าข้อสอบและตัวเลือกที่เหมาะสมกับนักเรียนแต่ละชั้นควรมีลักษณะอย่างไร ซึ่งเท่ากับว่าเป็นการส่งเสริมความเชี่ยวชาญของผู้สร้างข้อสอบให้สามารถระบุความสามารถนักเรียนที่ต้องการวัดได้แม่นยำมากขึ้น

5) เนื่องจากแบบสอบ ONET ของแต่ละปีการศึกษาเป็นแบบสอบคนละฉบับกัน และไม่มีหลักฐานเชิงประจักษ์เพื่อสนับสนุนความเป็นคู่ขนานกันของแบบสอบ (parallel test) หรือความสมมูลกันของแบบสอบ (equivalent test) จึงไม่สามารถเปรียบเทียบคะแนนสอบของแต่ละปีได้โดยตรง คะแนนสอบค่าเดียวกันจากต่างปีการศึกษาจะบ่งชี้ความสามารถนักเรียนไม่เหมือนกัน การทำแผนที่ภาวะสันนิษฐานจึงเป็นแนวทางหนึ่งที่สกัดสารสนเทศร่วมกันจากข้อสอบ ONET แต่ละข้อแล้วมาจัดกลุ่มตามระดับความซับซ้อนของข้อสอบและคำบรรยายความสามารถนักเรียนที่ใช้แก้ปัญหา

ข้อสอบที่มีความซับซ้อนแตกต่างกัน ซึ่งเมื่อพิจารณาคำบรรยายเหล่านี้จะพบว่าสามารถใช้บรรยายความสามารถนักเรียนในภาพรวมได้ดี แต่มีสารสนเทศเชิงวินิจฉัยไม่มากนัก

6) ข้อกำหนดลักษณะแบบสอบ ONET แต่ละปีการศึกษาแตกต่างกัน ทั้งในเรื่องของจำนวนข้อสอบที่วัดแต่ละสาระการเรียนรู้ไม่เท่ากัน และจำนวนมาตรฐานการเรียนรู้ที่วัดในแต่ละปีไม่เท่ากัน เมื่อพิจารณารายละเอียดของข้อกำหนดลักษณะแบบสอบ ONET ในแต่ละปีการศึกษา พบว่าแบบสอบ ONET ไม่ได้วัดครบทุกมาตรฐานการเรียนรู้ ซึ่งเท่ากับว่าแบบสอบ ONET ไม่สามารถให้สารสนเทศรายมาตรฐานการเรียนรู้ได้ เมื่อนำข้อกำหนดลักษณะแบบสอบดังกล่าวมาใช้สร้างแผนที่ภาวะสันนิษฐานจึงทำให้ไม่สามารถใช้บรรยายความสามารถนักเรียนรายมาตรฐานการเรียนรู้ด้วย

7) โมเดลที่ใช้ประมาณค่าความสามารถนักเรียนและค่าความยากข้อสอบในโครงการวิจัยนี้คือโมเดล MRCML จุดเด่นของโมเดลดังกล่าวคือสามารถออกแบบโมเดลการวัดแบบพหุมิติได้ด้วย โดยจะนำสารสนเทศจากข้อสอบ ONET ที่วัดสาระการเรียนรู้อื่นมาใช้ประมาณค่าความสามารถนักเรียนในสาระการเรียนรู้ที่สนใจผ่านความสัมพันธ์ระหว่างสาระการเรียนรู้ต่างๆ นั้นหมายความว่า ในทางทฤษฎีนั้นสาระการเรียนรู้วิทยาศาสตร์ที่ปรากฏในหลักสูตรแกนกลางฯ มีหลายด้านและสัมพันธ์กัน และในกรณีที่การทดสอบ ONET มีข้อสอบจำนวนไม่มากต่อหนึ่งฉบับแต่ต้องวัดให้ครอบคลุมทุกสาระการเรียนรู้ ก็จะประมาณค่าความสามารถนักเรียนของแต่ละสาระการเรียนรู้ได้ ผลการวิจัยแสดงให้เห็นว่าแม้ว่าข้อสอบ ONET ปีการศึกษา 2552 ที่มีจำนวนเพียง 37 ข้อ แต่มีสาระการเรียนรู้ที่มุ่งวัดถึง 8 สาระการเรียนรู้ โมเดล MRCML ก็สามารถประมาณค่าความสามารถนักเรียนรายสาระการเรียนรู้ และสามารถนำค่าประมาณความสามารถนักเรียนดังกล่าวไปใช้สร้างสเกลคะแนนแฝงและกำหนดคะแนนจุดตัดขั้นต้นรายสาระการเรียนรู้ได้อีกด้วย

8) ในกระบวนการกำหนดลักษณะโมเดล (model specification) ผู้เชี่ยวชาญได้ประเมินว่าข้อสอบข้อใดบ้างวัดสาระการเรียนรู้ที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี เนื่องจากสารดังกล่าวขาดหายไปจากกระบวนการออกแบบการวัด ซึ่งแม้ว่าจะสมมติให้ข้อสอบ ONET ทุกข้อวัดสารดังกล่าว แต่ผลจากการตรวจสอบความสอดคล้องกลมกลืนของโมเดลพบว่า ข้อสอบทุกข้อไม่ได้วัดสาระการเรียนรู้ที่ 8 โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสาระการเรียนรู้ที่ 8 กับสาระการเรียนรู้อื่นที่มีค่าต่ำหรือติดลบ และเมื่อผู้เชี่ยวชาญได้พิจารณาว่ามีข้อสอบบางข้อที่วัดสาระที่ 8 แต่ผลการพิจารณาดังกล่าวก็ยังไม่คงเส้นคงวา เพราะโมเดลการวัดที่พัฒนาขึ้นตามผลประเมินของผู้เชี่ยวชาญไม่สอดคล้องกลมกลืนกับข้อมูลคำตอบมากที่สุด (โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์ระหว่างสาระการเรียนรู้ที่ 8 กับสาระการเรียนรู้อื่นที่มีค่าต่ำหรือติดลบเช่นเดียวกัน) แสดงให้เห็นว่าความพยายามที่จะประมาณค่าความสามารถนักเรียนในสาระการเรียนรู้ที่ 8 ยังไม่ประสบความสำเร็จ ดังนั้นสเกลคะแนนและคะแนนจุดตัดขั้นต้นสาระการเรียนรู้ที่ 8 จึงยังไม่ถูกต้องแม่นยำเพียงพอ ผู้วิจัยจึงเสนอแนะแนวทางปรับปรุงแก้ไขการออกแบบโมเดลการวัดสาระการเรียนรู้ที่ 8 คือ นอกจากจะต้องให้ผู้เชี่ยวชาญพิจารณาข้อสอบที่คาดว่าจะวัดสารดังกล่าวใหม่อีกครั้งแล้ว ยังรวมถึงการ

ปรับปรุงคำบรรยายความสามารถนักเรียนในสาระการเรียนรู้ที่ 8 และการนำข้อสอบ ONET ปีการศึกษาอื่นมาพิจารณาเพิ่มขึ้นด้วย

ข้อเสนอแนะ

1) โครงการวิจัยนี้เสนอคะแนนจุดตัดขั้นต้นสองระดับคือคะแนนจุดตัดที่หนึ่งซึ่งกันระหว่างนักเรียนที่มีความสามารถขั้นต่ำกว่าขั้นพื้นฐานและขั้นพื้นฐาน และคะแนนจุดตัดที่สองซึ่งกันระหว่างความสามารถระดับขั้นพื้นฐานและขั้นสูง หลักการวิเคราะห์คะแนนจุดตัดที่เสนอนี้ขึ้นอยู่กับพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (item response theory) ซึ่งสามารถตีความคะแนนสอบ ONET ได้ว่านักเรียนที่มีคะแนนต่ำกว่าคะแนนจุดตัดจะมีแนวโน้มเฉลี่ยมากกว่า 50% ที่จะมีความสามารถตามคำบรรยายที่ระบุไว้ในแต่ละระดับความสามารถของแผนที่ภาวะสันนิษฐาน นักเรียนที่มีคะแนนห่างจากคะแนนจุดตัดมากขึ้นเท่าไร ก็จะมีแนวโน้มสูงกว่า 50% ขึ้นไปเรื่อยๆ ที่จะมีประสิทธิภาพตรงตามที่บรรยายไว้ สำหรับนักเรียนที่มีคะแนนใกล้เคียงกับคะแนนจุดตัดอาจมีข้อโต้แย้งว่าแนวโน้มที่ 50% อาจจะไม่เพียงพอเนื่องจากโดยธรรมชาติของการประมาณค่าความสามารถจะทำให้แนวโน้มดังกล่าวมีความคลาดเคลื่อนรวมอยู่ด้วย การตัดสินใจว่านักเรียนที่มีคะแนนต่ำกว่าจุดตัดสำคัญเพียงเล็กน้อยจึงจำเป็นต้องตรวจสอบมากยิ่งขึ้น โดยผู้วิจัยเสนอแนะว่าควรพิจารณาจากรูปแบบคำตอบ (response pattern) รายข้อเพื่อให้มั่นใจว่านักเรียนที่มีคะแนนใกล้เคียงกับคะแนนจุดตัดมีรูปแบบคำตอบแบบใด คำตอบของนักเรียนเหล่านั้นบ่งชี้อย่างชัดเจนหรือไม่ว่านักเรียนควรจะอยู่ในระดับขั้นที่สูงกว่า รูปแบบคำตอบที่ขาดหลักฐานเพียงพอและแสดงถึงความไม่คงเส้นคงวาของความสามารถนักเรียนที่มีต่อข้อสอบแต่ละข้อที่วัด จะนำไปสู่ข้อสรุปว่านักเรียนที่มีคะแนนใกล้เคียงกับคะแนนจุดตัดสมควรจะอยู่ในระดับความสามารถต่ำกว่า

คะแนนจุดตัดขั้นต้นนี้สามารถนำไปใช้กำหนดมาตรฐานความสามารถของแต่ละโรงเรียนได้เป็นอย่างดี ผู้วิจัยเสนอแนะว่าครูสามารถนำผลการประเมินการเรียนรู้ในชั้นเรียนมาเชื่อมต่อกับแผนที่ภาวะสันนิษฐานได้ โดยครูพิจารณาภาระงานแต่ละชิ้นว่าเหมาะสมกับความสามารถนักเรียนชั้นใด ผลงานของนักเรียนแต่ละชิ้นสะท้อนความสามารถขั้นใด รวมทั้งข้อสอบที่ครูสร้างขึ้นมุ่งวัดความสามารถนักเรียนระดับใดตามแผนที่ภาวะสันนิษฐานของโครงการวิจัยนี้ เมื่อครูเชื่อมโยงข้อมูลผลประเมินในชั้นเรียนเข้ากับผลประเมิน ONET ระดับชาติแล้วจะเห็นภาพที่ชัดเจนขึ้นว่า การประเมินการเรียนรู้ในชั้นเรียนนั้นมีทิศทางเกี่ยวกับการประเมินระดับชาติหรือไม่ ผลงานที่นักเรียนทำในระดับชั้นเรียนสอดคล้องกับความสามารถนักเรียนที่พิจารณาได้จากคะแนนจุดตัดหรือไม่

2) การตีความคะแนนสอบ ONET โดยเปรียบเทียบกับแผนที่ภาวะสันนิษฐานและคะแนนจุดตัดขั้นต้นเป็นการบรรยายแนวโน้มหรือโอกาสมากกว่าหรือน้อยกว่า 50% ที่นักเรียนจะมีประสิทธิภาพตรงตามที่กำหนด ผู้กำหนดคะแนนจุดตัดอาจต้องการเพิ่มความน่าจะเป็นให้สูงขึ้น เช่น ต้องการให้โอกาสขั้นต่ำที่นักเรียนมีความสามารถในแต่ละระดับตรงตามคำบรรยายใน

แผนที่ภาวะสันนิษฐานเป็น 80% ก็สามารถทำได้ โดยคำนวณค่าความน่าจะเป็นจากค่าความยากข้อสอบและค่าความสามารถนักเรียนที่ .80 แทนที่จะเป็น .50 คะแนนจุดตัดขั้นต้นชุดใหม่ที่จะเกิดขึ้นจึงเพิ่มสูงขึ้นจากเดิมด้วย ดังนั้นแนวทางการตีความคะแนนจุดตัดขั้นต้นที่หนึ่งในฐานะคะแนนการผ่านขั้นต่ำสุดจะแกร่งมากยิ่งขึ้น ผู้มีอำนาจตัดสินจะมั่นใจมากขึ้นว่านักเรียนที่มีคะแนนใกล้เคียงกับคะแนนจุดตัดชุดใหม่จะมีโอกาสอย่างน้อย 80% ที่มีอยู่ในขั้นความสามารถนั้นๆ

3) คะแนนจุดตัดขั้นต้นและแผนที่ภาวะสันนิษฐานสามารถให้สารสนเทศแบบสรุปรวมและแบบส่งเสริมการเรียนรู้ของนักเรียนได้ นักเรียนที่มีคะแนนสอบ ONET อยู่ในช่วงต่ำกว่าคะแนนจุดตัดที่หนึ่งจะประเมินความสามารถของตนเองชัดเจนมากขึ้นว่ารู้อะไรและทำอะไรได้บ้าง รวมทั้งขีดจำกัดต่างๆ ที่ทำให้นักเรียนไม่สามารถอยู่ในระดับขั้นที่สูงขึ้นได้ นั่นหมายความว่า นักเรียนจะทราบชัดเจนมากขึ้นว่านักเรียนไม่รู้อะไรและทำอะไรไม่ได้ด้วย เมื่อสารสนเทศนี้ถูกส่งกลับมาในระดับโรงเรียน ครูจะเข้าใจขีดจำกัดในการเรียนรู้ของนักเรียนแต่ละคน รวมทั้งประสิทธิภาพการสอนของครูที่อาจจะไม่ได้ส่งเสริมหรือไม่สอดคล้องกับแนวทางการทดสอบ ONET ด้วย ครูจะเข้าใจว่าถ้าจะส่งเสริมให้นักเรียนที่มีความสามารถอยู่ในขั้นต่ำกว่าขั้นพื้นฐานให้สูงขึ้นจะต้องให้นักเรียนเรียนรู้รู้อะไรและต้องฝึกให้นักเรียนมีทักษะอะไรเพิ่มขึ้น รวมทั้งในกรณีที่โรงเรียนมีนักเรียนจำนวนมากอยู่ในขั้นต่ำกว่าขั้นพื้นฐานก็จะมีแนวทางการส่งเสริมผลสัมฤทธิ์ทางการเรียนของนักเรียนที่ชัดเจนว่าในภาพรวมทั้งโรงเรียนจะต้องเน้นเรื่องใด คะแนนจุดตัดจึงทำหน้าที่สะท้อนมาตรฐานการศึกษาในระดับโรงเรียนว่าเป็นไปในทิศทางเกี่ยวกับการทดสอบระดับชาติหรือไม่ เช่นนักเรียนในโรงเรียนส่วนใหญ่มีผลการเรียนเฉลี่ยวิชาวิทยาศาสตร์สูง แต่เมื่อเปรียบเทียบกับคะแนนสอบ ONET พบว่านักเรียนมีคะแนนอยู่ในระดับขั้นพื้นฐานเท่านั้น ข้อค้นพบลักษณะนี้จะช่วยให้โรงเรียนปรับปรุงคุณภาพการตัดสินผลการเรียนของโรงเรียนให้อยู่ระดับเดียวกับการทดสอบระดับประเทศ

4) โครงการวิจัยที่ไม่ครอบคลุมขั้นตอนการพิจารณาตัดสินคะแนนจุดตัดจริงโดยผู้เชี่ยวชาญหรือผู้มีอำนาจตัดสิน กระบวนการที่จะเกิดขึ้นต่อจากโครงการวิจัยนี้คือ คณะผู้เชี่ยวชาญและผู้มีอำนาจตัดสินจะนำข้อมูลเกี่ยวกับข้อสอบ ONET คำบรรยายระดับความสามารถนักเรียนที่อยู่ในรูปของแผนที่ภาวะสันนิษฐาน และคะแนนจุดตัดขั้นต้นที่ผู้วิจัยเสนอแนะไปใช้ประกอบการตัดสินใจ ซึ่งผู้ควบคุมและดำเนินกระบวนการกำหนดคะแนนจุดตัดสามารถเลือกใช้วิธีกำหนดคะแนนจุดตัดอื่นได้ตามความเหมาะสม

เอกสารอ้างอิง

- กรมวิชาการ กระทรวงศึกษาธิการ (2545). *หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544*.
กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- กระทรวงศึกษาธิการ (2552). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*.
กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- วีรภัทร์ สุขศิริ (2557). การสร้างแผนที่ตัวแปรเชิงทฤษฎี: แนวทางการสร้างโมเดลความคิดในระบบ
การประเมินที่ถูกต้องและน่าเชื่อถือ. *วารสารวิจัยและพัฒนาหลักสูตร*, 4(1), 46-66.
- ศิริพันธ์ ตียะวงศ์สุวรรณ ญัณฐภรณ์ หลาวทอง และศิริชัย กาญจนวาสี (2556). การกำหนดมาตรฐาน
การประเมินการศึกษา: แนวคิด วิธีการและการปรับใช้. *วารสารวิทยาศาสตร์
สาขาสังคมศาสตร์*, 34(2), 225-240.
- สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (2552). *ตัวชี้วัดและสาระการเรียนรู้แกนกลางกลุ่ม
สาระการเรียนรู้วิทยาศาสตร์ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551*.
กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย
- สำนักงานเลขาธิการสภาการศึกษา (2552). ข้อเสนอการปฏิรูปการศึกษาในทศวรรษที่สอง (พ.ศ.
2552-2561). กรุงเทพมหานคร: พรักหวานกราฟฟิค.
- สำนักวิชาการและมาตรฐานการศึกษา กระทรวงศึกษาธิการ (2551). *ตัวชี้วัดและสาระการเรียนรู้
แกนกลาง กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน
พุทธศักราช 2551*. โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- Adams, R. J., & Khoo, S-T. (1996). *Quest: The interactive test analysis system*.
Camberwell, Australia: Australian Council for Educational Research
- Adams, R. J., Wilson, M., Wang, W. -C. (1997). The multidimensional random
coefficients multinomial logit model. *Applied Psychological Measurement*,
21(1), 1-23.
- Adams, R.J, Wu, M.L, & Wilson, M.R. (2012). ACER ConQuest version 3.0.1[computer
software]. Melbourne: ACER Press.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction
on Automatic Control*, 19(6), 716-723.
- American Educational Research Association, American Psychological Association, &
National Council on Measurement in Education (2014). *Standards for
educational and psychological testing*. Washington, DC: American Educational
Research Association.

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Angoff, W. A. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Atkinson, D. (2012). Legal issues and considerations for standard setting in professional licensure and certification examinations. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 503-534). New York, NY: Routledge.
- Baker, K. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, MD: University of Maryland.
- Bejar, I. I., Braun, H. L., & Tannenbaum, R. J. (2007). A perspective, progressive, and predictive approach to standard setting. In R. W. Lissitz (Ed.), *Assessing and modeling cognitive development in school* (pp. 1-30). Maple Grove, MN: JAM.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56(1), 137-172.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model*. Mahwah, NJ: Erlbaum.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440.
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, 6(1), 25-53.
- Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4(1), 87-100.
- Brown, N., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychological Review*, 23(2), 221-234.
- Bunch, M. B. (2012). Practical issues in standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 415-438). New York, NY: Routledge.
- Bunch, M. B. (2013). Construct maps for the road ahead. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 171-173.
- Cizek, G. J. (1996). Setting passing scores. *Educational Measurement: Issues and Practice*, 15(2), 20-31.

- Cizek, G. J. (2012). An introduction to contemporary standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 3-14). New York, NY: Routledge.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Davis-Becker, S. (2013). Construct maps: Do they make the unclear clear? *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 174-176.
- Draney, K., & Kennedy, C. (2010). The standard setting criterion-mapping method. In C. Harsch, H. A. Pant, & O. Koller (Eds.). *Calibrating standards-based assessment task for English as a first foreign language: Vol. 2. Standard-setting procedures in Germany* (pp 75-88). Berlin, Germany: Waxmann Verlag GmbH.
- Draney, K., Kennedy, C., Moore, L., & Morell, L. (2010). Procedural standard-setting issues. In C. Harsch, H. A. Pant, & O. Koller (Eds.). *Calibrating standards-based assessment task for English as a first foreign language: Vol. 2. Standard-setting procedures in Germany* (pp 89-122). Berlin, Germany: Waxmann Verlag GmbH.
- Draney, K., & Wilson, M. (2011). Selecting cut scores with a composite of item types: The Construct Mapping procedure. *Journal of Applied Measurement*, 12(3), 298-309.
- Ebel, R. L. (1972). *Essentials of educational measurement*. Englewood Cliffs, NJ: Printice-Hall.
- Embretson, S. E. (2007). Construct validity: A Universal validity system or just another test evaluation procedure? *Educational Research*, 36(8), 449-455.
- Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed framework. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 79-106). New York, NY: Routledge.
- Fischer, G. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3– 26.
- Fischer, G. (1995). Derivations of the Rasch model. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15-38). New York: Springer-Verlag.

- Guttman, L. (1944). A basis of scaling qualitative data. *American Sociological Review*, 9(2), 139-150.
- Guttman, L. (1950). The basis of scalogram analysis. In S. A. Stouffer, L. A. Guttman, F. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Studies in social psychology in World War II: Vol. 4. Measurement and prediction*. (pp. 60-90). Princeton: Princeton University Press.
- Guttman, L. (1954). An outline of some new methodology for social research. *Public Opinion Quarterly*, 18(4), 395-404.
- Haertel, E. H. (2002). Standard setting as a participatory process: Implications for validation of standards-based accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 16-22.
- Hambleton, R. M. (1998). Setting performance standards on achievement tests: Meeting the requirement of Title I. In L. N. Hansche (Ed.), *Handbook for the development of performance standards* (pp.87-114). Washington, DC: Council of Chief State School Officers.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24(4), 355-366.
- Hambleton, R. H., Swaminathan, H., & Rogers, H. J. (1991). *Foundations of item response theory*. Newbury Park, CA: SAGE.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433-470). Westport, CT: Praeger.
- Harsch, C., Pant, H. A., & Koller, O. (2010). *Calibrating standards-based assessment task for English as a first foreign language: Vol. 2. Standard-setting procedures in Germany*. Berlin, Germany: Waxmann Verlag GmbH.
- Jaeger, R. M. (1982). An interactive structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis*, 4(4), 461-475.
- Jaeger, R. M. (1995). Setting standards for complex performances: An iterative, judgmental policy-capturing strategy. *Educational Measurement: Issues and Practice*, 14(4), 16-20.

- Jaeger, R. M., & Mills, C. M. (2001). An integrated judgment procedure for setting standards on complex, large-scale assessment. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 313-338). Mahwah, NJ: Lawrence Erlbaum.
- Kaftandjieva, F. (2010). *Methods for setting cut scores in criterion-references achievement tests: A comparative analysis of six recent methods with an application to test of reading in EFL*. Arnhem, The Netherlands: CITO.
- Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425-461.
- Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53-81). Mahwah, NJ: Lawrence Erlbaum.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education/Praeger.
- Kane, M. T., & Tannenbaum, R. J. (2013). The role of construct maps in standard setting. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 177-180.
- Katz, I. R., Tannenbaum, R. J., & Kannan, P. (2009). Virtual standard setting. *CLEAR Exam Review*, 20(2), 19-27.
- Kingston, N. M., Kahl, S. R., Sweeney, K. P., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Kingston, N. M., & Tiemann, G. C. (2012). Setting performance standards on complex assessment: The body of work ahead method. G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp.201-224). New York, NY: Routledge.
- Kingston, N. M., Tiemann, G. C., & Loughran, J. T. (2013). Commentary on "Construct Maps as a foundation for standard setting." *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 185-188.

- Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531-578). Westport, CT: American Council on Education/Praeger.
- Lehrer, R., Kim, M-J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In A. Maloney, J. Confrey & K. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 31-60). Charlotte, NC: Information Age Publishers.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). *Standard setting: A Bookmark approach*. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225-253). New York, NY: Routledge.
- Lin, T. S. & Dayton, C. M. (1997). Model-selection information criteria for nonnested latent class models. *Journal of Educational and Behavioral Statistics*, 22, 249-264.
- Linacre, J. M. (1990). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Linacre, J. M. (1994). Sample size and item calibration [or person measure] stability. *Rasch Measurement Transactions*, 7(4), 328.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Loomis, S. C., & Bourque, M. L. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 175-218). Mahwah, NJ: Lawrence Erlbaum.
- Luecht, R. M. (2013). Assessment engineering task model maps, task models and templates as a new way to develop and implement test specifications. *Journal of Applied Testing Technology*, 14, 1-38.
- MacCann, R. G., & Stanley, G. (2006). The use of Rasch modeling to improve standard setting. *Practical Assessment Research & Evaluation*, 11(2), 1-11.

- Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. Charlotte, NC: Information Age Publishing.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Masters, J. S. (2010). *A comparison of traditional test blueprinting and item development to assessment engineering in a licensure context*. Retrieved from https://libres.uncg.edu/ir/uncg/f/Masters_uncg_0154D_10377.pdf
- McClarty, K. L. (2013). Construct maps: A tool to organize validity evidence. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 185-188.
- Mehrens, W. A., & Cizek, G. J. (2012). Standard setting for decision making: Classifications, consequences, and the common good. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 33-46). New York, NY: Routledge.
- Mellenbergh, G. J. (2011). *A Conceptual introduction to psychometrics: Development, analysis, and application of psychological and educational tests*. The Hague: Eleven international publishing.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379-416.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-67.
- Michell, J. (2008). Is psychometricians pathological science? *Measurement: Interdisciplinary Research and Perspectives*, 6(1), 7-24.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum.

- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser (Eds). Washington, DC: National Academy Press.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.
- O'Malley, K., Keng, L., & Miles, J. (2012). From Z to A: Using validity evidence to set performance standards. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 301-322). New York, NY: Routledge.
- OECD (2007). *Executive summary PISA 2006: Science competencies for tomorrow's world*. Paris: OECD Publishing.
- OECD (2010). *PISA 2009 results: Executive summary*. Paris: OECD Publishing.
- OECD (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Paris: OECD Publishing.
- Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practice*, 27(4), 15-29.
- Phillips, G. W. (2010). *International benchmarking: State education performance standards*. Washington, DC: American Institutes for Research.
- Phillips, S. E. (2012). Legal issues for standard setting in K-12 educational contexts. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 535-569). New York, NY: Routledge.
- Plake, B. S., & Hambleton, R. K. (2001). The analytical judgment method for setting standards on complex performance assessments. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 283-312). Mahwah, NJ: Lawrence Erlbaum.
- Plake, B. S., Huff, K., & Reshester, R. (2010). Evidence-centered assessment design as a foundation for achievement-level descriptor development and for standard setting. *Applied Measurement in Education*, 23, 342-357.
- Putnam, S. E., Pence, P., & Jaeger, R. M. (1995). A multi-stage dominant profile method for setting cut scores on complex performance assessments. *Applied Measurement in Education*, 8(1), 57-83.

- Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago Press).
- Reckase, M. D. (1998). *Analysis of methods for collecting test-based judgments*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Reckase, M. D. (2000). A survey and evaluation of recently developed procedures for setting standards on educational tests. In M. L. Bourque, & S. Byrd (Eds.), *Student Performance Standards on the National Assessment of Educational Progress: Affirmations and Improvements* (pp. 43-69). Washington, DC: National Assessment Governing Board.
- Raymond, M. R. (2016). Job analysis, practice analysis and the content credentialing examinations. In S. Lane, M. R. Raymond & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed., pp. 144-164). New York, NY: Routledge.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- Schulz, E. M. (2013). What is essential in standard setting and construct maps? Commentary on Adam E. Wyse's "Construct Maps as a Foundation for Standard Setting. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 192-194.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Shih, C-K., Chang, H-L., & Cheng, M-H. (2013). Introducing multidimensional Rasch model in measuring traffic police officers' behavior. *Proceeding of the Eastern Asia Society for Transportation Studies*, 9, 1-20.
- Skaggs, G. (2013). Coming full circle in standard setting: A commentary on Wyse. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 195-197.
- Skorupski, W. P. (2012). Understanding the cognitive processes of standard setting panelists. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 135-148). New York, NY: Routledge.
- Smith, R. M., Schumacker, R. E., & Busch, M. J. (1995). *Using item mean squares to evaluate fit to the Rasch model*. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stone, M., & Yumoto, F. (2004). The effect of sample size for estimating Rasch/IRT parameters with dichotomous items. *Journal of Applied Measurement*, 5(1), 48-61.
- Tannenbaum, R. J., & Katz, I. R. (2013). Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol 3. Testing and assessment in school psychology and education* (pp. 455-477). Washington, DC: American Psychology Association.
- van der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In van der Linden & Hambleton, R. K. (Eds.), *Handbook of modern item response theory* (pp. 1-28), New York: Springer.
- Wang, N. (2003). Use of the Rasch IRT model in standard setting: An item mapping approach. *Journal of Educational Measurement*, 40, 231-253.
- Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.
- Wilson, M. (1992). The ordered partition model: An extension of the partial credit model. *Applied Psychological Measurement*, 16, 309-325.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Wilson, M. (2015). *Applications in educational monitoring and practice: The logic of standard-setting* [PowerPoint slides]. Retrieved from http://www.uv.uio.no/ceemo/english/conferences/Standard-setting2015/presentations/standard_wilson.ppt.
- Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, 60(2), 181-198.
- Wilson, M., Bejar, I., Scalise, K., Templin, J., William, D., & Irribarra, D. T. (2012). Perspectives on methodological issues. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 67-141). New York, NY: Springer.
- Wilson, M., & Boeck, D. P. (2004). Descriptive and explanatory item response models. In P. D. Boeck, & M. Wilson (Eds.), *Explanatory item response models* (pp. 43-74). New York, NY: Springer.

- Wilson, M., & Draney, K. (2002). A technique for setting standards and maintaining them over the time. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), *Measurement and multivariate analysis* (pp. 325-332). New York, NY: Springer-Verlag.
- Wilson, M., & Wang, W. C. (1995). Complex composites: Issues that arise in combining different models of assessment. *Applied Psychological Measurement*, 19(1), 51-71.
- Wright, B. D. (1984). Additivity in psychological measurement. In E. E. Roskam (Ed.), *Measurement and personality measurement* (pp. 101-111). Amsterdam, North-Holland: Elsevier Science.
- Wright, B. D. (1992). What is the “right” test length? *Rasch Measurement Transactions*, 6(1), 205.
- Wright, B. D., Mead, R. J., & Bell, S. R. (1980). *BICAL: Calibrating items with the Rasch model*. Chicago, IL: Research Memorandum.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Wu, M. L., Adams, R. L., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modeling software manual*. Melbourne: ACER Press.
- Wyse, A. E. (2013). Construct map as a foundation for standard setting. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 139-170.
- Wyse, A. E. (2014). Construct maps in standard setting: A response and some additional thoughts. *Measurement: Interdisciplinary Research and Perspectives*, 12, 62-68.
- Wyse, A. E., Bunch, M. B., Deville, C., & Viger, S. G. (2013). A body of work standard-setting method with construct maps. *Educational and Psychological Measurement*, 74(2), 236-262.
- Zieky, M. J. (2012). So much has changed: An historical overview of setting cut scores. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 15-32). New York, NY: Routledge.
- Zieky, M. J., Perie, M., & Livingston, S. A. (2008). *Cutscores: A manual for setting standards of performance in educational and occupational tests*. Princeton, NJ: Educational Testing Service.

Zhou, J. (2009). *A review of assessment engineering principles with select application to the certified public accountant examination: Technical report*. Durham, NC: The American Institute of Certified Public Accountants.

ภาคผนวก

ตารางที่ A1 คำบรรยายและส่วนประกอบของมาตรฐานการเรียนรู้วิทยาศาสตร์ตามหลักสูตร
แกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551

มาตรฐาน	มาตรฐานเนื้อหา	มาตรฐานการปฏิบัติ
ว 1.1	เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้างและหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน	มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต
ว 1.2	เข้าใจกระบวนการและความสำคัญของการถ่ายทอดลักษณะทางพันธุกรรม วิวัฒนาการของสิ่งมีชีวิต ความหลากหลายทางชีวภาพ การใช้เทคโนโลยีชีวภาพที่มีผลกระทบต่อมนุษย์และสิ่งแวดล้อม	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์
ว 2.1	เข้าใจสิ่งแวดล้อมในท้องถิ่น ความสัมพันธ์ระหว่างสิ่งแวดล้อมกับสิ่งมีชีวิต ความสัมพันธ์ระหว่างสิ่งมีชีวิตต่างๆ ในระบบนิเวศ	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์
ว 2.2	เข้าใจความสำคัญของทรัพยากรธรรมชาติ การใช้ทรัพยากรธรรมชาติในระดับท้องถิ่น ประเทศ และโลก	นำความรู้ไปใช้ในการจัดการทรัพยากรธรรมชาติ และสิ่งแวดล้อมในท้องถิ่นอย่างยั่งยืน
ว 3.1	เข้าใจสมบัติของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์
ว 3.2	เข้าใจหลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์
ว 4.1	เข้าใจธรรมชาติของแรงแม่เหล็กไฟฟ้า แรงโน้มถ่วง และแรงนิวเคลียร์	มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์อย่างถูกต้องและมีคุณธรรม
ว 4.2	เข้าใจลักษณะการเคลื่อนที่แบบต่างๆ ของวัตถุในธรรมชาติ	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์
ว 5.1	เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน ปฏิสัมพันธ์ระหว่างสารและพลังงาน ผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม	มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ ไปใช้ประโยชน์
ว 6.1	เข้าใจกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลก ความสัมพันธ์ของกระบวนการต่างๆ ที่มีผลต่อการเปลี่ยนแปลงภูมิอากาศ ภูมิประเทศและสัณฐานของโลก	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์
ว 7.1	เข้าใจวิวัฒนาการของระบบสุริยะ กาแล็กซีและเอกภพ การปฏิสัมพันธ์ภายในระบบสุริยะและผลต่อสิ่งมีชีวิตบนโลก	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ การสื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์
ว 7.2	เข้าใจความสำคัญของเทคโนโลยีอวกาศที่นำมาใช้ในการสำรวจอวกาศและทรัพยากรธรรมชาติ ด้านการเกษตรและการสื่อสาร	มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างมีคุณธรรมต่อชีวิตและสิ่งแวดล้อม
ว 8.1		ใช้กระบวนการทางวิทยาศาสตร์และจิตวิทยาศาสตร์ในการสืบเสาะหาความรู้ การแก้ปัญหา รู้ว่าปรากฏการณ์ทางธรรมชาติที่เกิดขึ้นส่วนใหญ่มีรูปแบบที่แน่นอน สามารถอธิบายและตรวจสอบได้ ภายใต้ข้อมูลและเครื่องมือที่มีอยู่ในช่วงเวลานั้นๆ เข้าใจว่าวิทยาศาสตร์ เทคโนโลยี สังคม และสิ่งแวดล้อมมีความเกี่ยวข้องสัมพันธ์กัน

ตารางที่ A2 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ข้อสอบ ONET ที่วัดมาตรฐาน ว 8.1

ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน		ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน		ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน	
		วัด ว 8.1	ไม่วัด ว 8.1			วัด ว 8.1	ไม่วัด ว 8.1			วัด ว 8.1	ไม่วัด ว 8.1
ปีการศึกษา 2550											
1	4	0	1	16	4	0	1	31*	4	1	0
2	4	0	1	17	4	.25	.75	32	4	0	1
3	4	0	1	18	4	.50	.50	33*	4	1	0
4	4	.50	.50	19	4	.50	.50	34	4	.25	.75
5	4	.50	.50	20	4	.25	.75	35	4	.25	.75
6	4	0	1	21	4	.50	.50	36	4	0	1
7	4	.50	.5	22	4	.50	.50	37	4	.25	.75
8*	4	.75	.25	23	4	.50	.50	38	4	.50	.50
9	4	.50	.50	24	4	.50	.50	39	4	.25	.75
10*	4	.75	.25	25	4	.25	.75	40	4	.25	.75
11	4	.50	.50	26*	4	1	0	เกณฑ์: ค่าดัชนีความสอดคล้องมากกว่า .50 ข้อสอบที่ผ่านเกณฑ์มีจำนวน 5 ข้อ ได้แก่ ข้อ 8, 10, 26, 31 และ 33			
12	4	0	1	27	4	.50	.50				
13	4	.25	.75	28	4	.50	.50				
14	4	0	1	29	4	.50	.50				
15	4	.25	.75	30	4	0	1				
ปีการศึกษา 2551											
1	7	.14	.86	16	5	0	1	31*	4	1	0
2	6	0	1	17	5	0	1	32	4	0	1
3*	7	1	0	18*	5	1	0	33	4	0	1
4	6	0	1	19	5	0	1	34	4	0	1
5	5	0	1	20	5	0	1	35	4	0	1
6	5	.20	.80	21*	4	.75	.25	36	3	.33	.67
7	5	0	1	22	4	.50	.50	37	4	.75	.25
8	4	.25	.75	23	5	.25	.75	38	4	0	1
9	6	0	1	24*	4	1	0	39	3	0	1
10	5	0	1	25	3	.33	.67	40	3	0	1
11	4	.25	.75	26	3	.33	.67	เกณฑ์: ค่าดัชนีความสอดคล้องมากกว่า .50 ข้อสอบที่ผ่านเกณฑ์มีจำนวน 9 ข้อ ได้แก่ ข้อ 3, 14, 18, 21, 24, 27, 28, 30 และ 31			
12	3	0	1	27*	4	1	0				
13	5	0	1	28*	4	1	0				
14*	5	1	0	29	4	.50	.50				
15	5	0	1	30*	4	1	0				

ตารางที่ A2 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ข้อสอบ ONET ที่วัดมาตรฐาน ว 8.1 (ต่อ)

ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน		ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน		ข้อที่	จำนวน ผู้เชี่ยวชาญ	ผลประเมิน	
		วัด ว 8.1	ไม่วัด ว 8.1			วัด ว 8.1	ไม่วัด ว 8.1			วัด ว 8.1	ไม่วัด ว 8.1
ปีการศึกษา 2552											
1	5	0	1	16	4	.25	.75	31	3	0	1
2	5	0	1	17	5	.20	.80	32	3	0	1
3*	6	1	0	18	3	.33	.67	33	3	0	1
4*	4	1	0	19	3	.33	.67	34	3	0	1
5	5	0	1	20	4	.25	.75	35	4	0	1
6	5	0	1	21	4	0	1	36	4	0	1
7*	5	1	0	22	5	0	1	37	4	0	1
8	5	0	1	23	3	0	1	เกณฑ์: ค่าดัชนีความสอดคล้องมากกว่า .50 ข้อสอบที่ผ่านเกณฑ์มีจำนวน 8 ข้อ ได้แก่ ข้อ 3, 4, 7, 10, 11, 13, 15 และ 27			
9	4	.25	.75	24	3	0	1				
10*	5	.80	.20	25	3	0	1				
11*	4	1	0	26	4	0	1				
12	4	0	1	27*	3	1	0				
13*	5	1	0	28	4	0	1				
14	5	.20	.80	29	4	0	1				
15*	5	.80	.20	30	4	0	1				

ตารางที่ A3 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ความสอดคล้องระหว่างความสามารถนักเรียนใช้ทำข้อสอบกับตัวชี้วัดความสามารถ

ข้อที่	มาตรฐาน และ ตัวชี้วัดที่ เกี่ยวข้อง ²	ผลประเมิน ¹		ข้อที่	มาตรฐาน และ ตัวชี้วัดที่ เกี่ยวข้อง ²	ผลประเมิน ¹		ข้อที่	มาตรฐาน และ ตัวชี้วัดที่ เกี่ยวข้อง ²	ผลประเมิน ¹	
		สอดคล้อง	ไม่ สอดคล้อง			สอดคล้อง	ไม่ สอดคล้อง			สอดคล้อง	ไม่ สอดคล้อง
ปีการศึกษา 2550											
1	ว 1.1 5/2	1	0	16	ว 3.1 6/4	.75	.25	31 ³	ว 6.1 6/1	1	0
2	ว 1.1 5/3	1	0	17	ว 3.1 6/5	1	0	32	ว 6.1 6/2	1	0
3	ว 1.1 5/4	.75	.25	18	ว 3.2 6/1	1	0	33 ³	ว 6.1 4/2	1	0
4	ว 1.1 6/3	1	0	19	ว 3.2 6.2	1	0	34	ว 6.1 5/3	.75	.25
5	ว 1.1 6/3	1	0	20	ว 4.1 5/1	1	0	35	ว 6.1 5/4	1	0
6	ว 1.1 6/2	1	0	21	ว 4.1 5/3	1	0	36	ว 7.2 6/1	1	0
7	ว 1.2 5.2	1	0	22	ว 4.2 5/1	1	0	37	ว 7.1 4/1	1	0
8 ³	ว 2.1 6/1	1	0	23	ว 5.1 4/2	1	0	38	ว 7.1 4/1	1	0
9	ว 2.1 6/2	1	0	24	ว 5.1 4/6	.75	.25	39	ว 7.1 5/1	1	0
10 ³	ว 3.1 5/1	1	0	25	ว 5.1 5/4	1	0	40	ว 7.1 5/1	.75	.25
11	ว 3.1 5/1	1	0	26 ³	ว 5.1 5/1	1	0				
12	ว 3.1 6/2	.75	.25	27	ว 5.1 6/4	1	0				
13	ว 3.1 6/3	1	0	28	ว 5.1 6/3	.75	.25				
14	ว 3.1 6/4	1	0	29	ว 5.1 6/5	1	0				
15	ว 3.1 6/5	1	0	30	ว 6.1 6/1	1	0				
ปีการศึกษา 2551											
1	ว 1.1 5/3	1	0	16	ว 3.1 6/4	1	0	31 ³	ว 6.1 5/3	.75	.25
2	ว 1.2 5/2	1	0	17	ว 3.1 6/4	.80	.20	32	ว 6.1 5/4	1	0
3 ³	ว 1.1 6/3	1	0	18 ³	ว 3.2 6/1+2	1	0	6/1			
4	ว 1.1 6/2	1	0	19	ว 4.1 5/1	1	0	33	ว 6.1 5/3+4	1	0
5	ว 1.2 5/2	1	0	20	ว 4.1 5/3	1	0	34	ว 7.1 4/1	1	0
6	ว 1.2 5/2	1	0	21 ³	ว 4.2 5/1	1	0	35	ว 7.1 6/1	.75	.25
7	ว 1.2 5/5	.60	.40	22	ว 5.1 4/3	1	0	36	ว 7.1 6/1	.67	.33
8	ว 2.1 6/1+3	1	0	23	ว 5.1 4/3	.60	.40	37	ว 7.1 6/1	1	0
9	ว 2.1 6/2	1	0	24 ³	ว 5.1 5/1	.75	.25	38	ว 7.2 6/1	1	0
10	ว 2.1 6/2	1	0	25	ว 5.1 5/2	1	0	39	ว 7.2 6/1	1	0
11	ว 3.1 5/1	1	0	26	ว 5.1 6/1	1	0	40	ว 7.2 6/1	1	0
12	ว 3.1 6/1	1	0	27 ³	ว 5.1 6/1	1	0				
13	ว 3.1 6/3	1	0	28 ³	ว 6.1 6/1	1	0				
14 ³	ว 3.1 6/2+3	1	0	29	ว 6.1 4/2	1	0				
15	ว 3.1 6/5	1	0	30 ³	ว 6.1 4/2	1	0				

ตารางที่ A3 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ความสอดคล้องระหว่างความสามารถนักเรียนใช้ทำข้อสอบกับตัวชี้วัดความสามารถ
(ต่อ)

ข้อที่	มาตรฐาน	ผลประเมิน ¹		ข้อที่	มาตรฐาน	ผลประเมิน ¹		ข้อที่	มาตรฐาน	ผลประเมิน ¹	
	และ ตัวชี้วัดที่ เกี่ยวข้อง ²	สอดคล้อง	ไม่ สอดคล้อง		และ ตัวชี้วัดที่ เกี่ยวข้อง ²	สอดคล้อง	ไม่ สอดคล้อง		และ ตัวชี้วัดที่ เกี่ยวข้อง ²	สอดคล้อง	ไม่ สอดคล้อง
ปีการศึกษา 2552											
1	ว 2.2 6/4	1	0	16	ว 3.1 5/1	.75	.25	31	ว 7.1 4/1	1	0
2	ว 1.1 6/3	1	0	17	ว 4.1 5/4	1	0	32	ว 7.1 6/1	1	0
3 ³	ว 1.1 4/2	1	0	18	ว 4.2 5/1	1	0	33	ว 2.1 6/2	1	0
4 ³	ว 1.1 4/2	1	0	19	ว 5.1 6/5	1	0	34	ว 4.1 5/4	1	0
5	ว 1.2 5/2	1	0	20	ว 5.1 4/4	1	0	35	ว 5.1 5/2	1	0
6	ว 1.2 5/5	1	0	21	ว 5.1 4/4	1	0	36	ว 7.1 4/1	1	0
7 ³	ว 4.1 5/4	1	0	22	ว 5.1 4/3	1	0	37	ว 1.1 5/4	.75	.25
8	ว 3.1 6/3	1	0	23	ว 6.1 6/3	.67	.33				
9	ว 3.1 6/2	.75	.25	24	ว 6.1 6/1	.67	.33				
10 ³	ว 3.1 5/1	.80	.20	25	ว 1.1 5/2	1	0				
11 ³	ว 3.1 6/1	1	0	26	ว 1.1 6/2	1	0				
12	ว 3.1 5/2	.75	.25	27 ³	ว 1.1 4/2	1	0				
13 ³	ว 3.1 5/1	.80	.20	28	ว 7.1 4/1	1	0				
14	ว 4.1 5/4	1	0	29	ว 7.1 6/1	1	0				
15 ³	ว 3.1 6/5	1	0	30	ว 6.1 4/2	.75	.25				

หมายเหตุ: ¹ ผู้เชี่ยวชาญมีจำนวนเท่ากับที่แสดงในตารางที่ A2

² แสดงมาตรฐานที่วัด ระดับชั้นและตัวชี้วัดของระดับชั้นนั้นโดยมีลำดับตามหลักสูตรแกนกลางฯ พ.ศ. 2551 เช่น
ว 1.1 5/1 หมายถึงข้อสอบข้อนั้นวัดมาตรฐาน ว 1.1 ระดับชั้น ป.5 ตัวชี้วัดลำดับที่ 1

³ ข้อสอบที่วัดมาตรฐาน 8.1 ร่วมกับมาตรฐานอื่นไม่ได้วิเคราะห์ความสัมพันธ์กับตัวบ่งชี้ของมาตรฐาน 8.1

ตารางที่ A4 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ระดับความยากข้อสอบ ONET

ข้อ	ผลประเมินระดับความยากข้อสอบ				ข้อ	ผลประเมินระดับความยากข้อสอบ			
	ง่าย	ปานกลาง	ค่อนข้างยาก	ยาก		ง่าย	ปานกลาง	ค่อนข้างยาก	ยาก
ปีการศึกษา 2550 ¹									
1	.40	.60	-	-	21	.20	.60	.20	-
2	.40	.60	-	-	22	.80	.20	-	-
3	1	-	-	-	23	.40	.60	-	-
4	-	.40	-	.60	24	.80	-	.20	-
5	-	.20	.60	.20	25	-	.40	.60	-
6	.40	.60	-	-	26	-	.20	.80	-
7	.80	-	-	.20	27	-	.80	-	.20
8	.20	.20	.60	-	28	.40	.60	-	-
9	.20	.20	.60	-	29	.20	.20	.60	-
10	-	.60	.40	-	30	1	-	-	-
11	-	.60	-	.40	31	-	.20	.80	-
12	1	-	-	-	32	.80	.20	-	-
13	.40	.60	-	-	33	.20	.20	.60	-
14	.60	-	.40	-	34	.80	-	.20	-
15	-	.80	-	.20	35	.60	.40	-	-
16	.40	.60	-	-	36	1	-	-	-
17	.80	-	.20	-	37	.40	.60	-	-
18	.20	.80	-	-	38	.20	.20	.60	-
19	.20	.60	.20	-	39	.60	.20	.20	-
20	-	.60	.20	.20	40	-	.20	-	.60
ปีการศึกษา 2551 ¹									
1	-	.29	.71	-	11	-	.75	.25	-
2	-	1	-	-	12	1	-	-	-
3	-	.14	.86	-	13	1	-	-	-
4	1	-	-	-	14	-	-	-	1
5	1	-	-	-	15	-	1	-	-
6	-	1	-	-	16	-	-	-	1
7	-	.75	-	.25	17	-	1	-	-
8	-	-	.25	.75	18	-	.20	.60	.20
9	1	-	-	-	19	1	-	-	-
10	-	.20	.60	.20	20	-	1	-	-

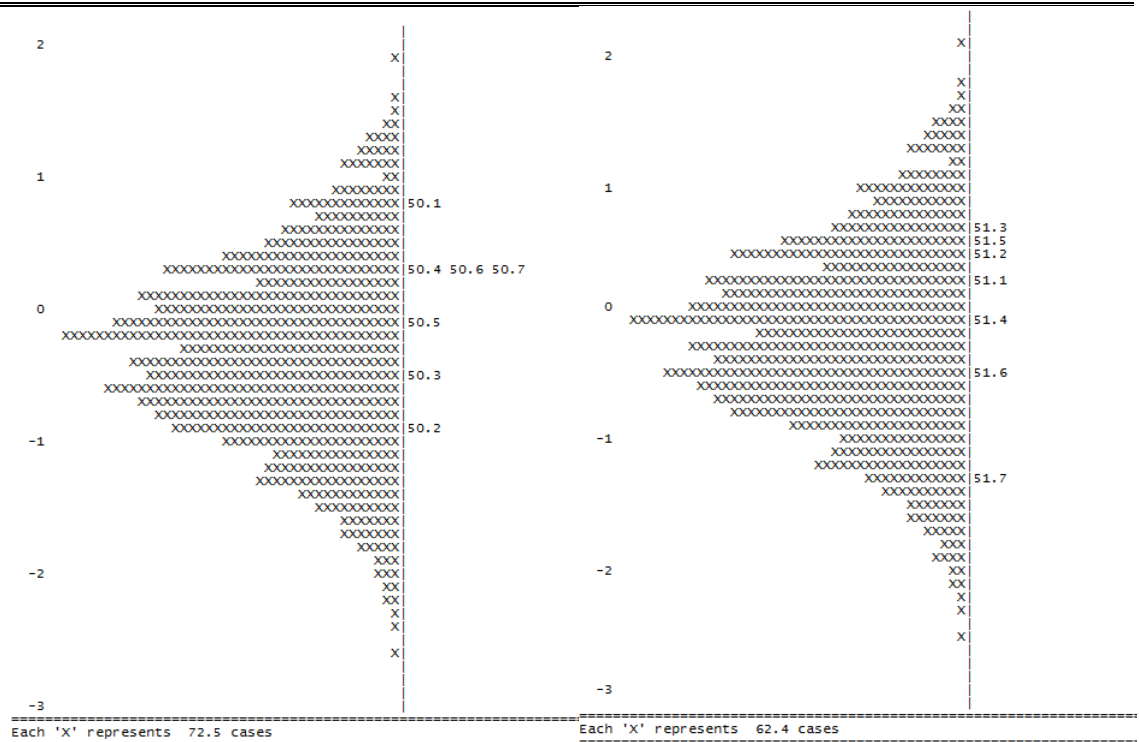
ตารางที่ A4 ค่าดัชนีความคล้ายคลึง (similarity index) ของผลประเมินโดยผู้เชี่ยวชาญเกี่ยวกับ
ระดับความยากข้อสอบ ONET (ต่อ)

ข้อ	ผลประเมินระดับความยากข้อสอบ				ข้อ	ผลประเมินระดับความยากข้อสอบ			
	ง่าย	ปานกลาง	ค่อนข้างยาก	ยาก		ง่าย	ปานกลาง	ค่อนข้างยาก	ยาก
ปีการศึกษา 2551 ¹									
21	-	.75	.25	-	31	-	1	-	-
22	1	-	-	-	32	-	.25	.75	-
23	1	-	-	-	33	.20	.60	.20	-
24	-	1	-	-	34	1	-	-	-
25	-	1	-	-	35	-	.20	.60	.20
26	.67	.33	-	-	36	.33	.67	-	-
27	-	1	-	-	37	.40	.60	-	-
28	-	.75	.25	-	38	1	-	-	-
29	.25	.75	-	-	39	.40	.60	-	-
30	.25	.75	-	-	40	.75	.25	-	-
ปีการศึกษา 2552 ¹									
1	1	-	-	-	21	.25	.75	-	-
2	1	-	-	-	22	.40	.60	-	-
3	-	-	.17	.83	23	.33	.67	-	-
4	-	-	.25	.75	24	.67	.33	-	-
5	1	-	-	-	25	-	1	-	-
6	-	-	1	-	26	.25	.75	-	-
7	-	1	-	-	27	-	-	.67	.33
8	1	-	-	-	28	1	-	-	-
9	.75	.25	-	-	29	1	-	-	-
10	.60	.40	-	-	30	.75	.25	-	-
11	.75	.25	-	-	31	-	.75	.25	-
12	.40	.60	-	-	32	1	-	-	-
13	.20	.60	.20	-	33	-	.33	.67	-
14	-	1	-	-	34	-	.25	.75	-
15	-	.60	.40	-	35	-	.25	.75	-
16	.40	.60	-	-	36	-	-	.75	.25
17	-	.40	.60	-	37	-	-	.75	.25
18	1	-	-	-	เกณฑ์: ข้อสอบที่มีค่าดัชนีความสอดคล้องมากกว่า .50 ถือว่ามีความยากตรงกับระดับที่กำหนด				
19	.75	.25	-	-					
20	.25	.75	-	-					

หมายเหตุ: ¹ จำนวนผู้เชี่ยวชาญที่พิจารณาข้อสอบแต่ละข้อของปีการศึกษา 2550 เท่ากับ 5 คน ของปีการศึกษา 2551 ระหว่าง 3-7 คน และปีการศึกษา 2552 ระหว่าง 3-6 คน

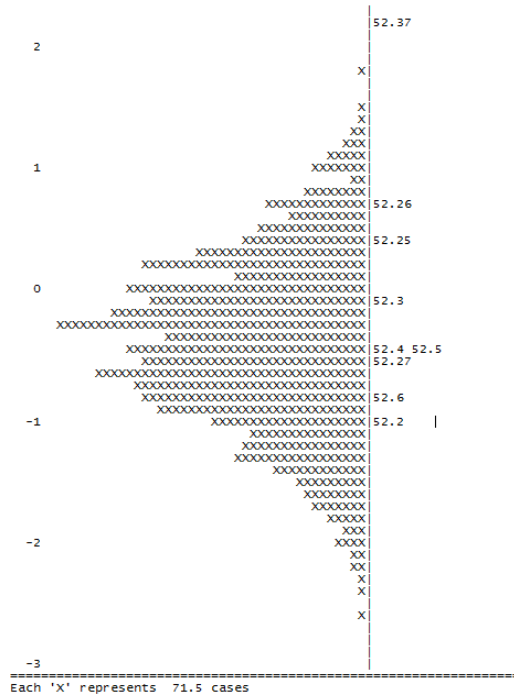
ภาพที่ A5 ผลวิเคราะห์ความสอดคล้องกลมกลืนระหว่างโมเดลการวัดกับข้อมูลคำตอบเพื่อ
ตรวจสอบคุณภาพแผนที่ภาวะสันนิษฐานเชิงทฤษฎีความสามารถทางวิทยาศาสตร์

A5.1 สารการเรียนรู้สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

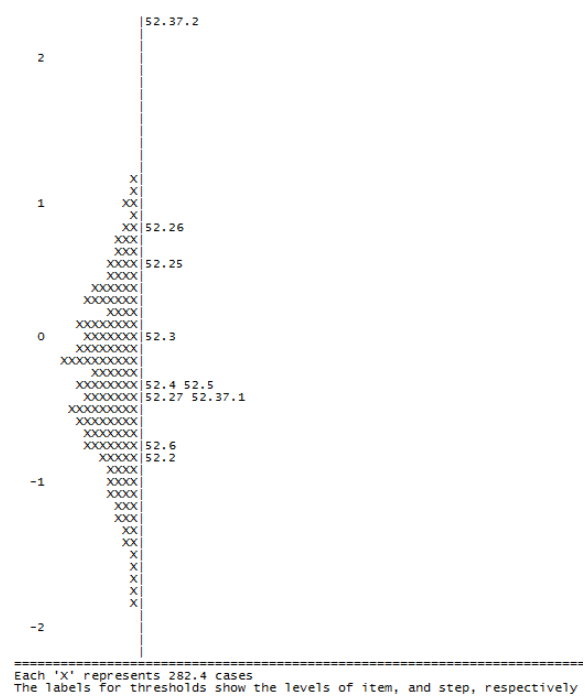


(a) ปีการศึกษา 2550

(b) ปีการศึกษา 2551

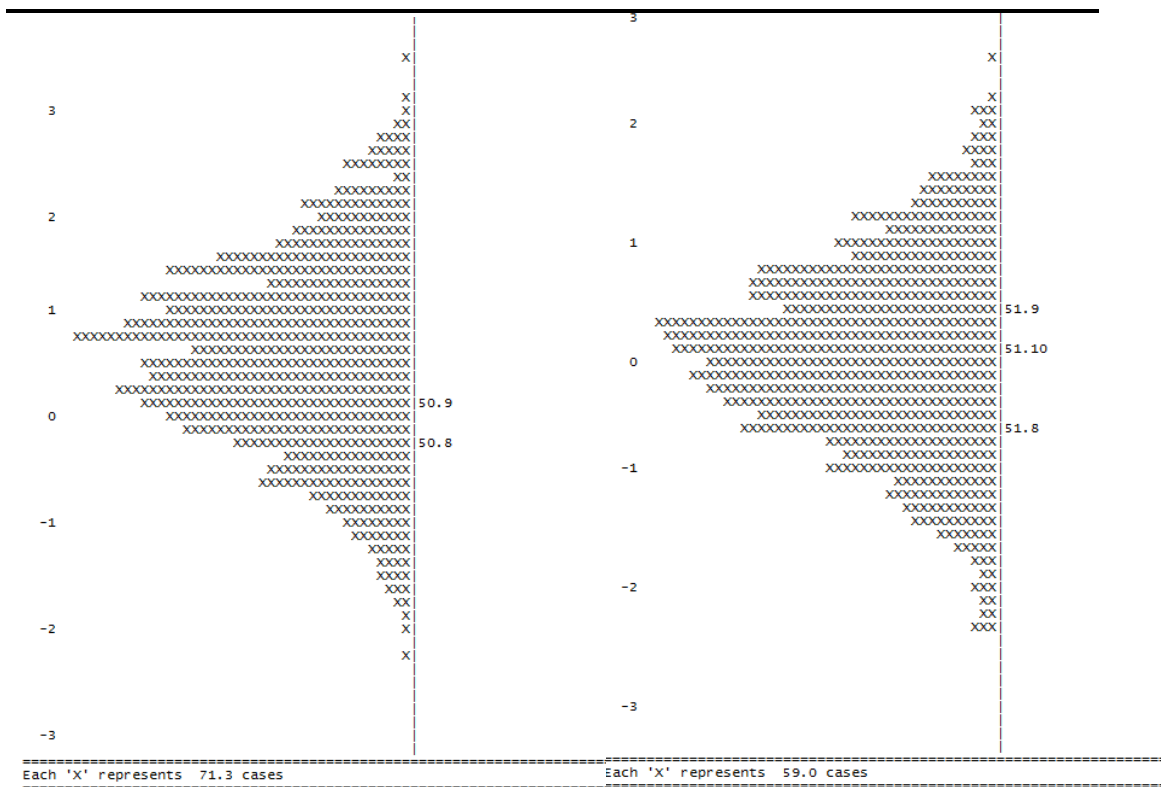


(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)



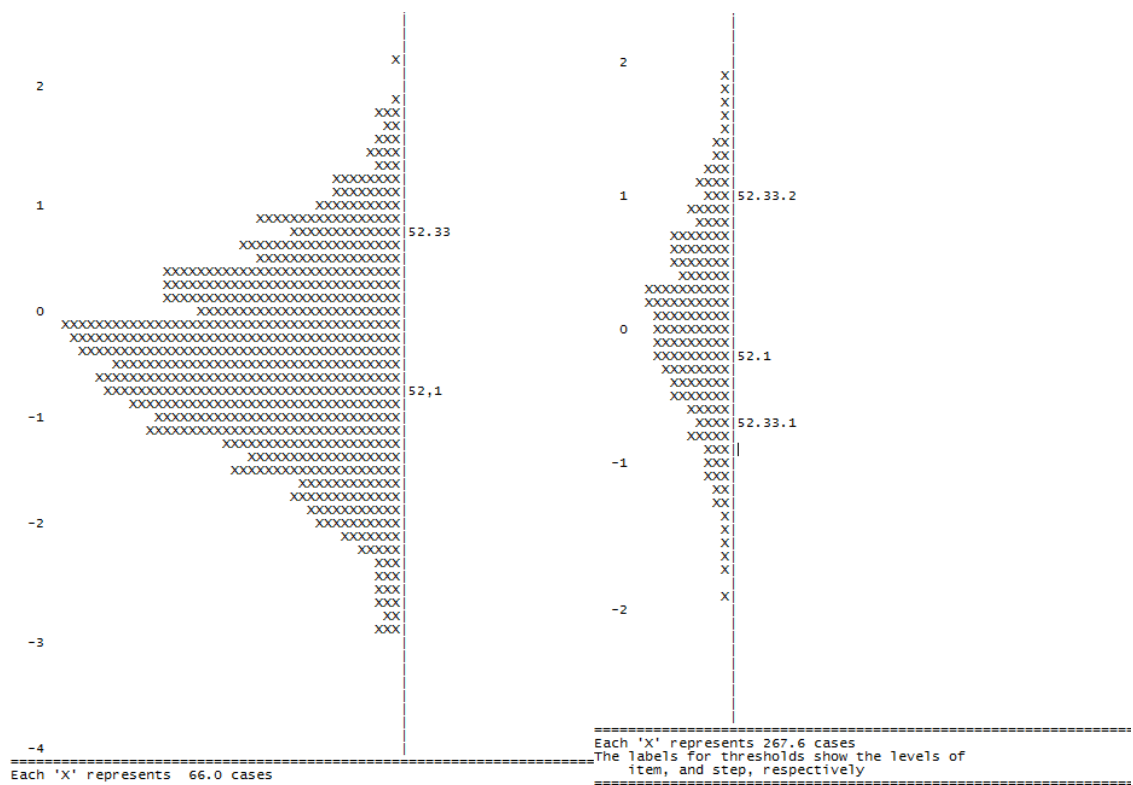
(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

A5.2 ตารางการเรียนรู้ชีวิตกับสิ่งแวดล้อม



(a) ปีการศึกษา 2550

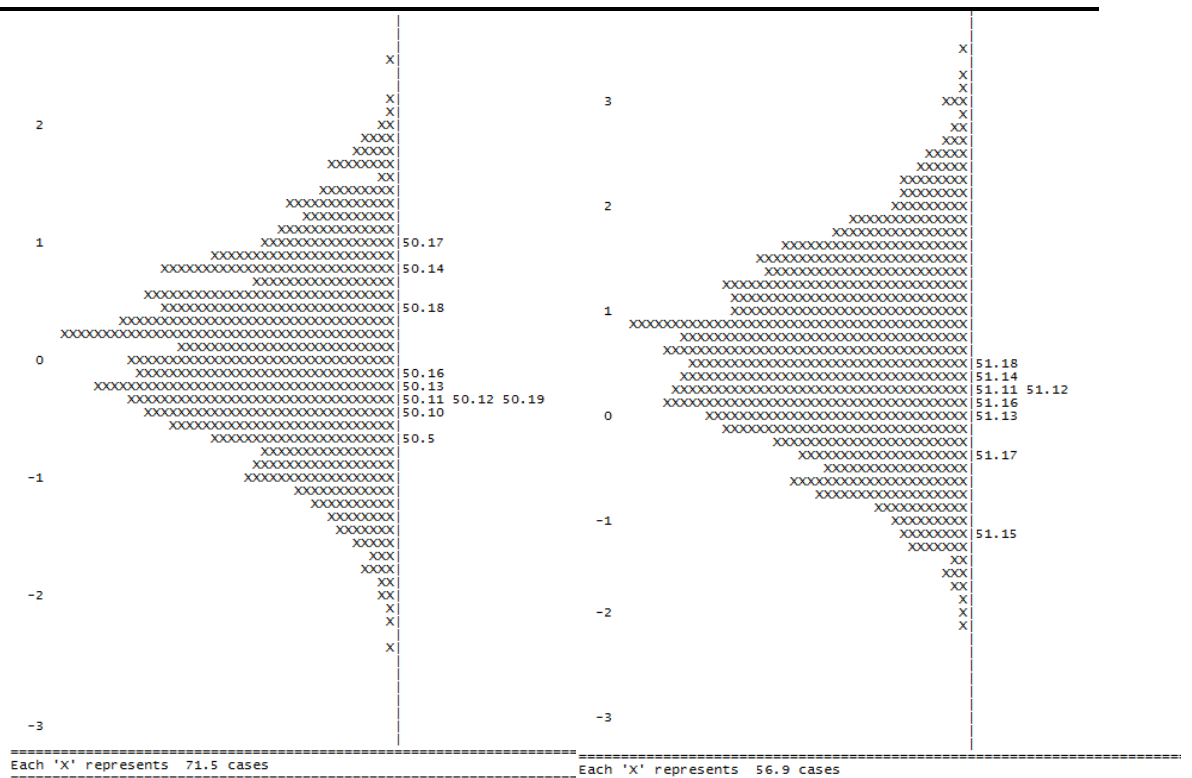
(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)

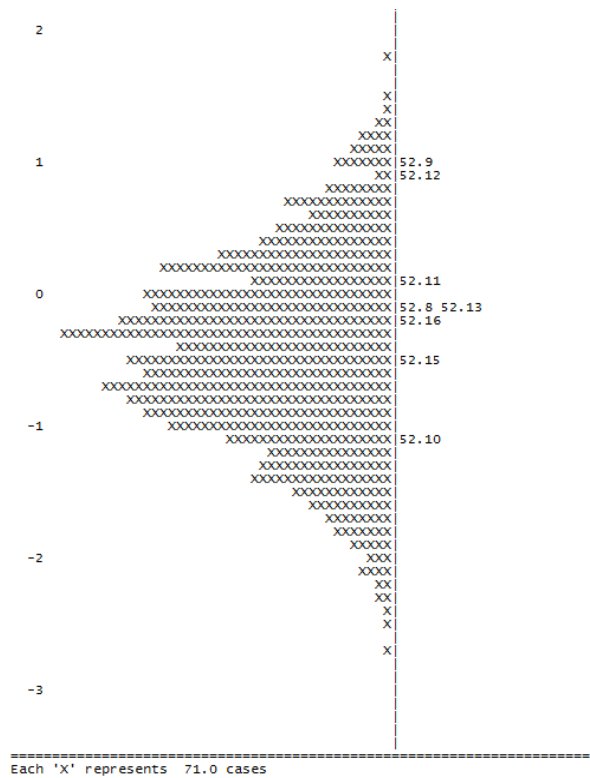
(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

A5.3 ตารางการเรียนรู้สารและสมบัติของสาร



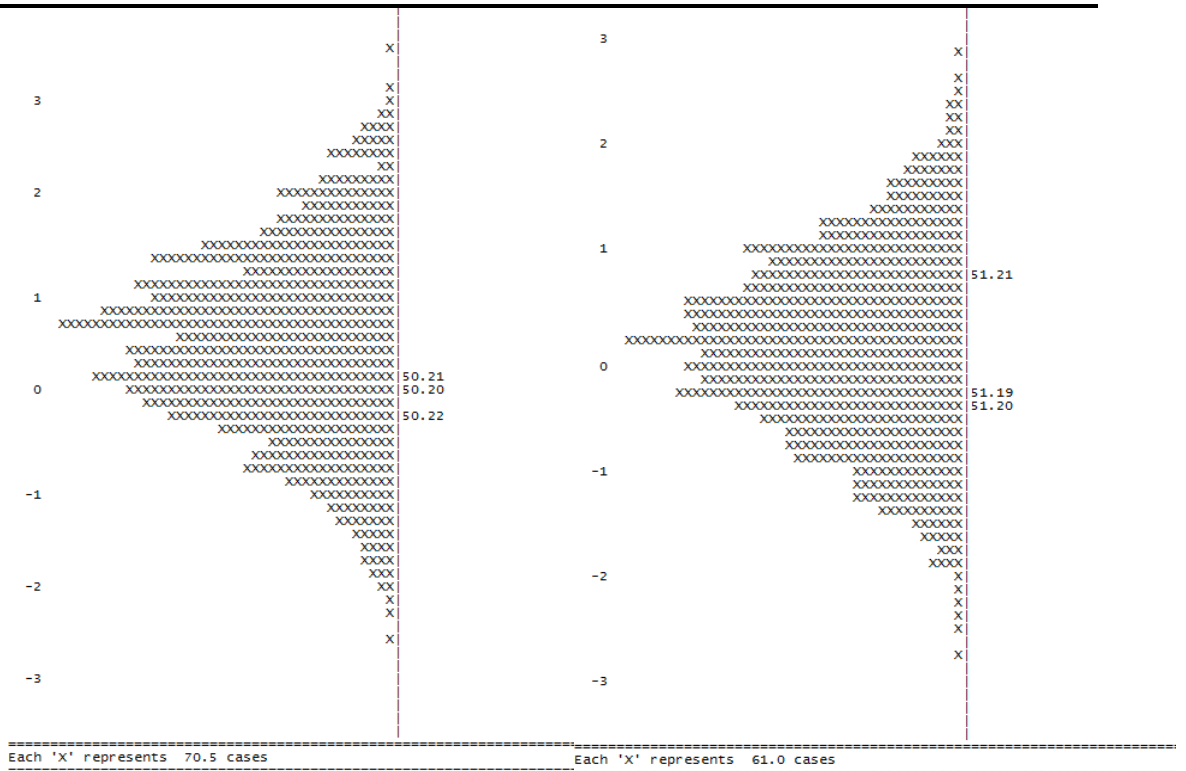
(a) ปีการศึกษา 2550

(b) ปีการศึกษา 2551



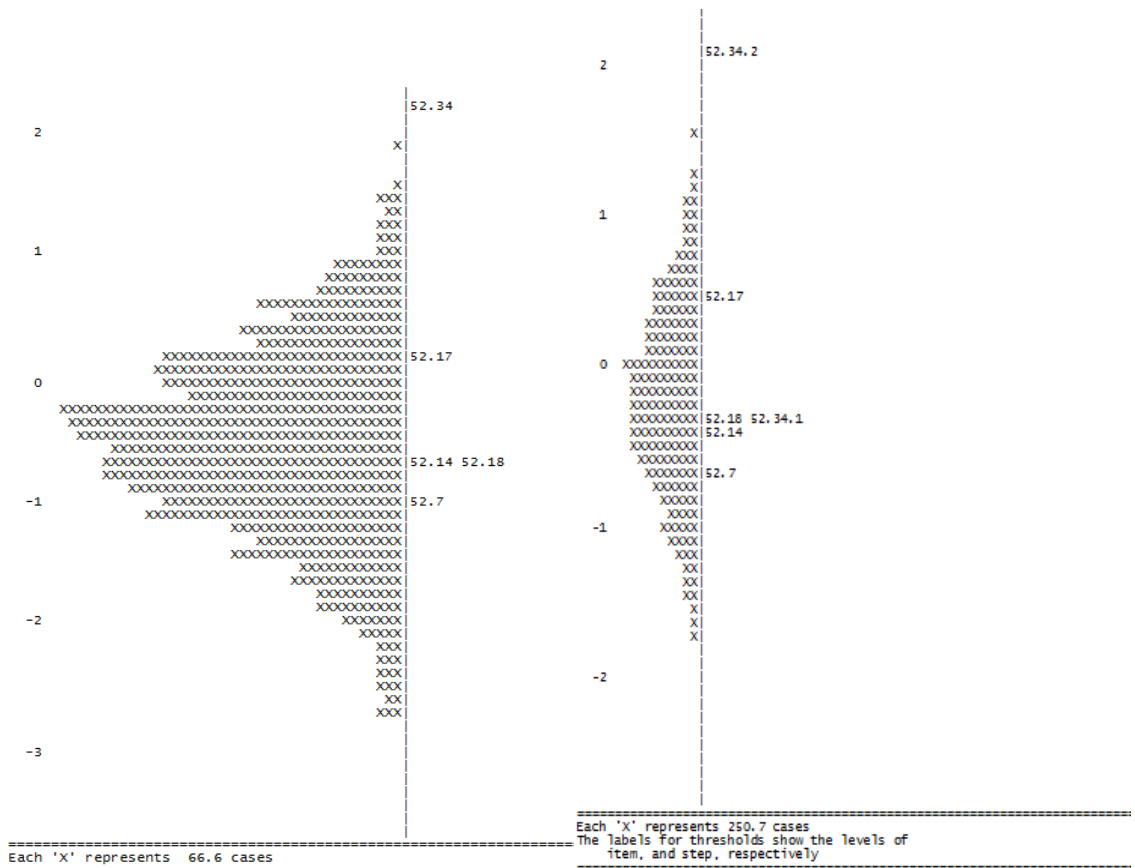
(c) ปีการศึกษา 2552

A5.4 ตารางการเรียนรู้แรงและการเคลื่อนที่



(a) ปีการศึกษา 2550

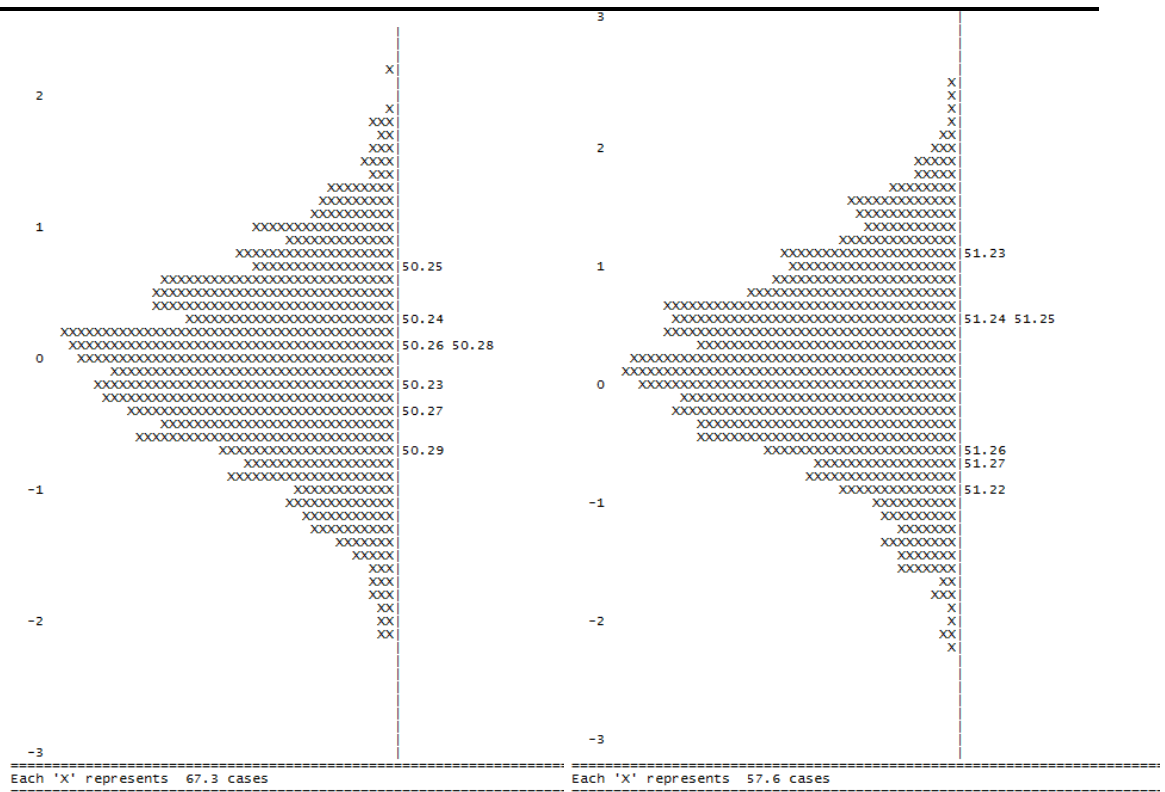
(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)

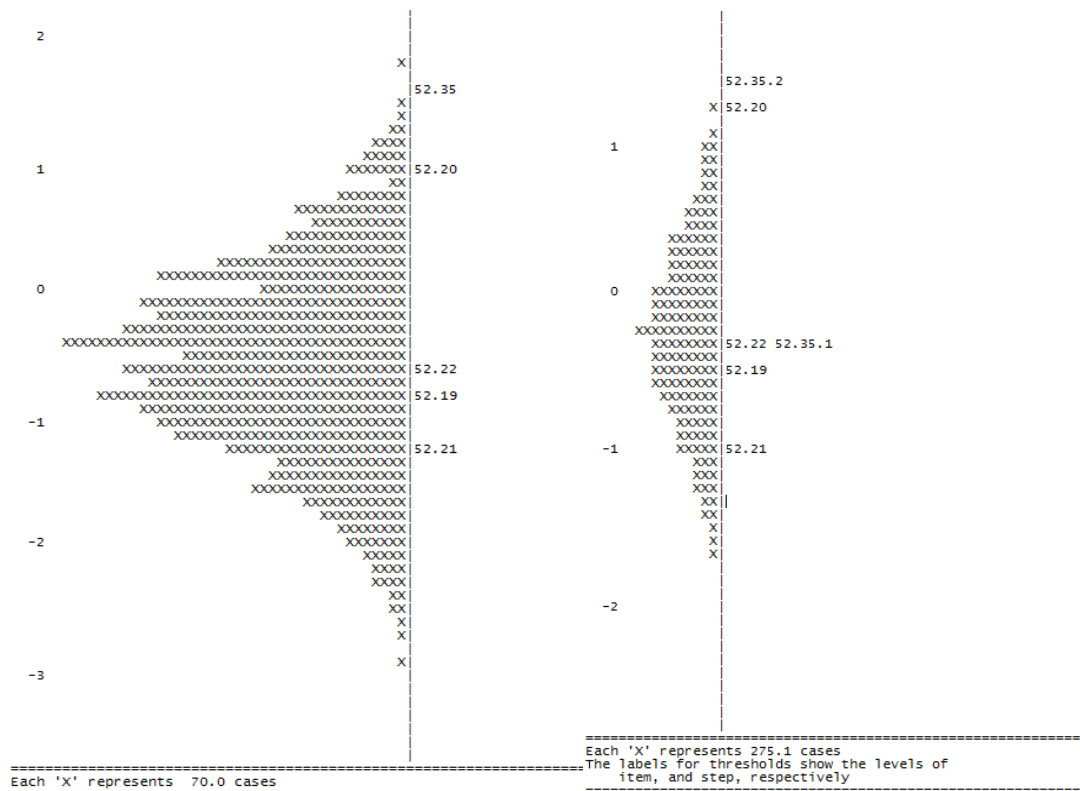
(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

A5.5 ตารางการเรียนรู้พลังงาน



(a) ปีการศึกษา 2550

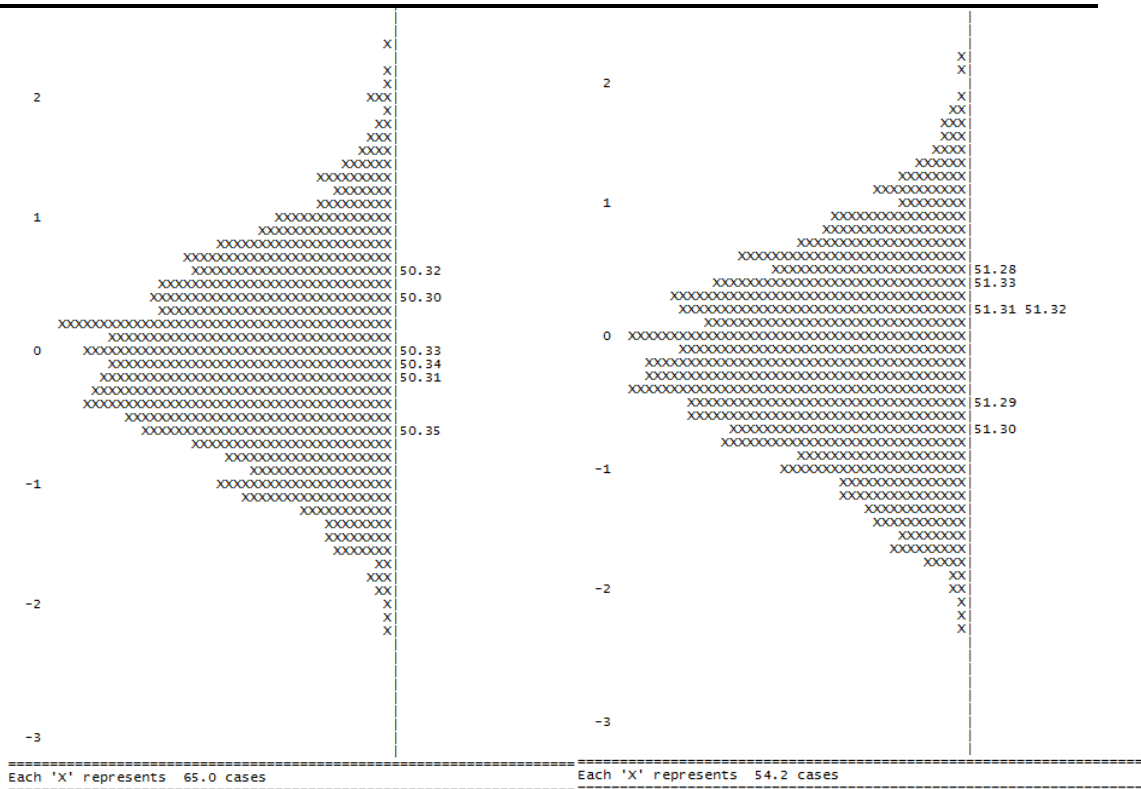
(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)

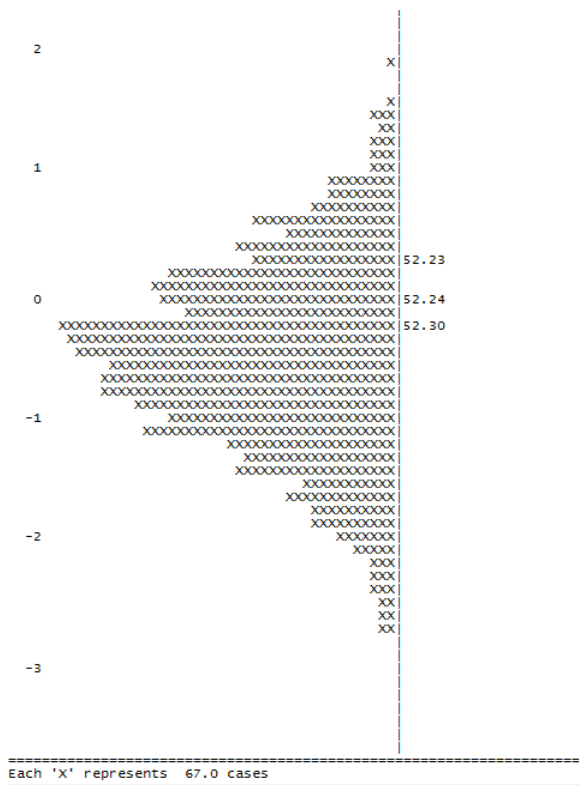
(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

A5.6 ตารางการเรียนรู้กระบวนการเปลี่ยนแปลงของโลก



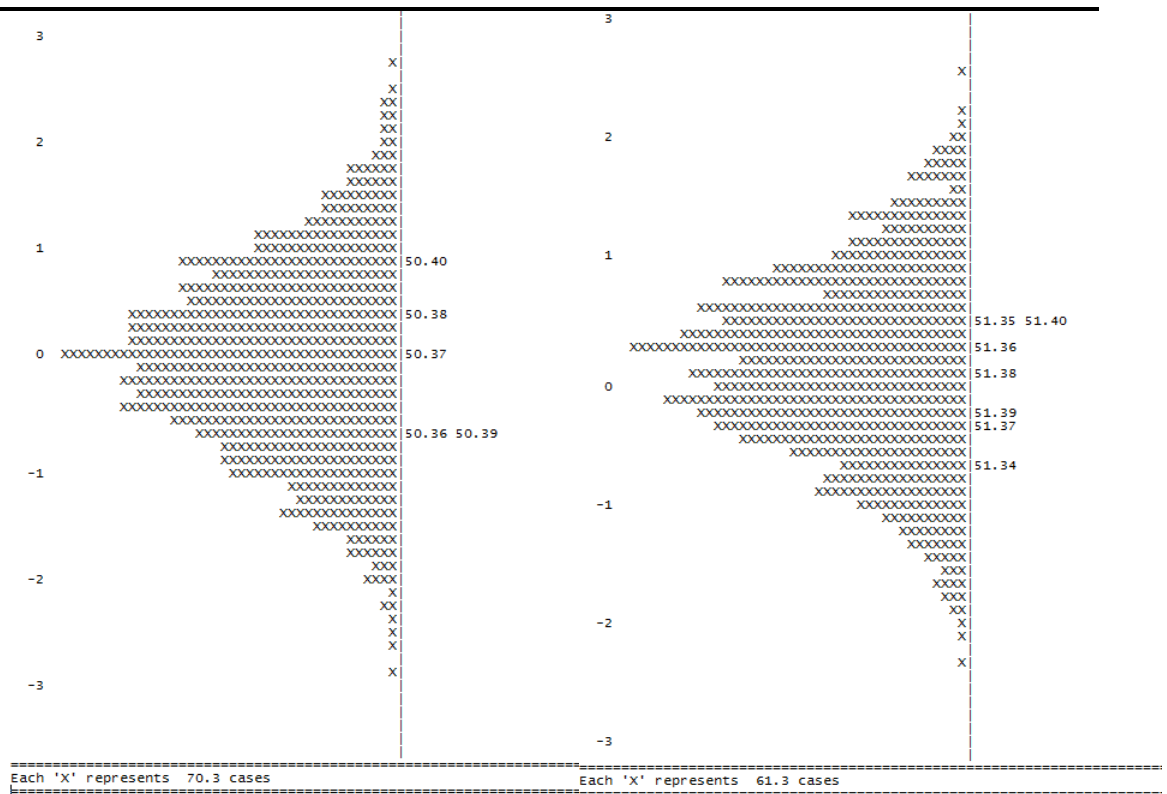
(a) ปีการศึกษา 2550

(b) ปีการศึกษา 2551



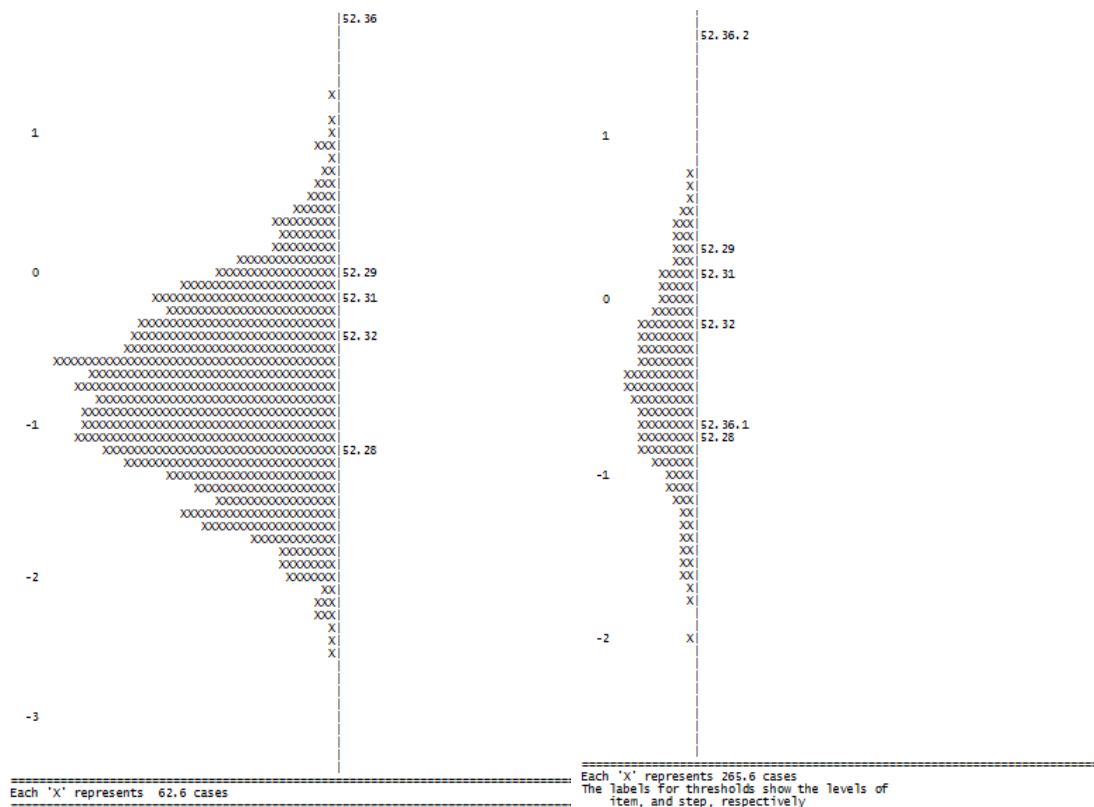
(c) ปีการศึกษา 2552

A5.7 ตารางการเรียนรู้ดาราศาสตร์และอวกาศ



(a) ปีการศึกษา 2550

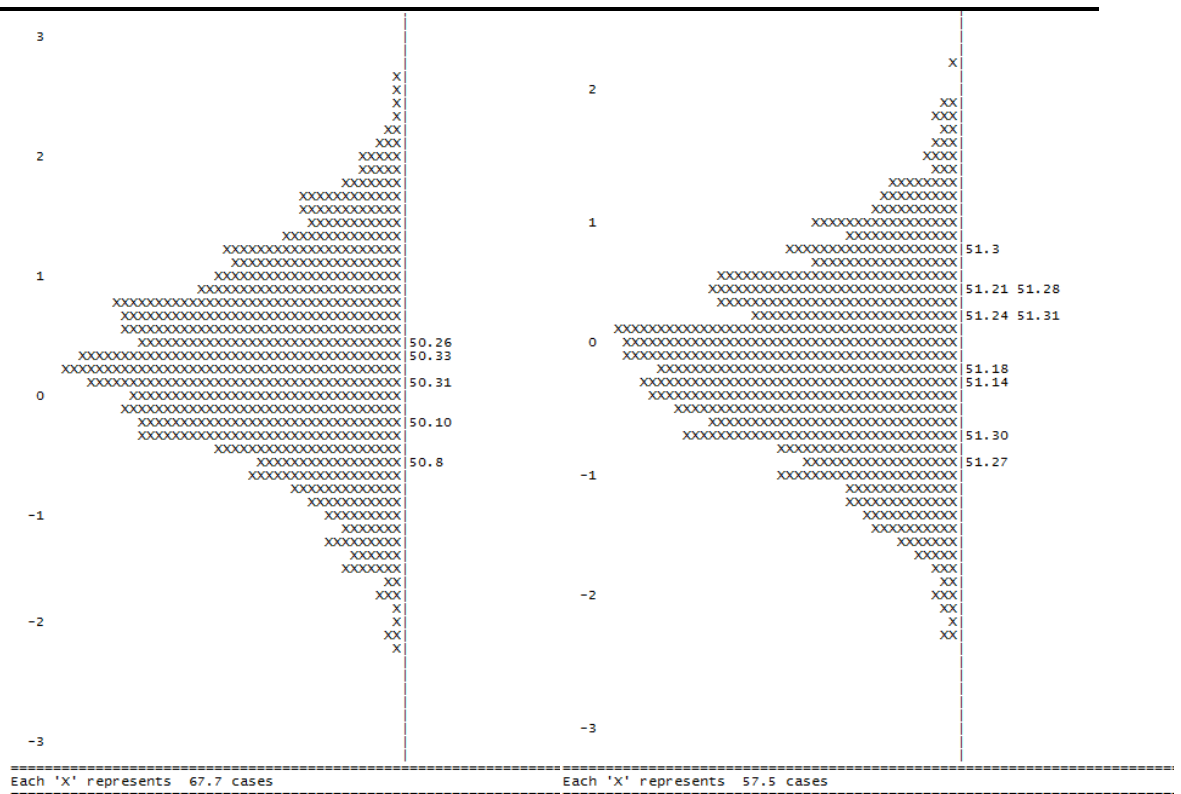
(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)

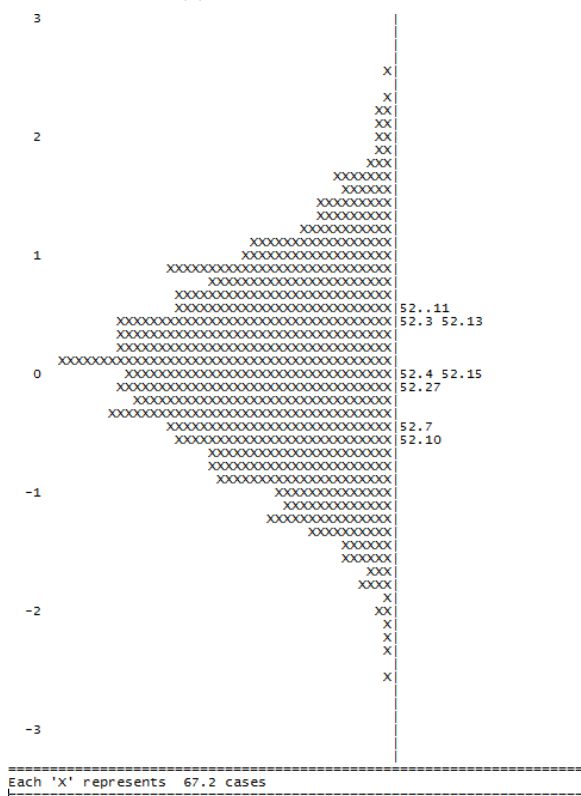
(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

A5.8 ตารางการเรียนรู้ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี



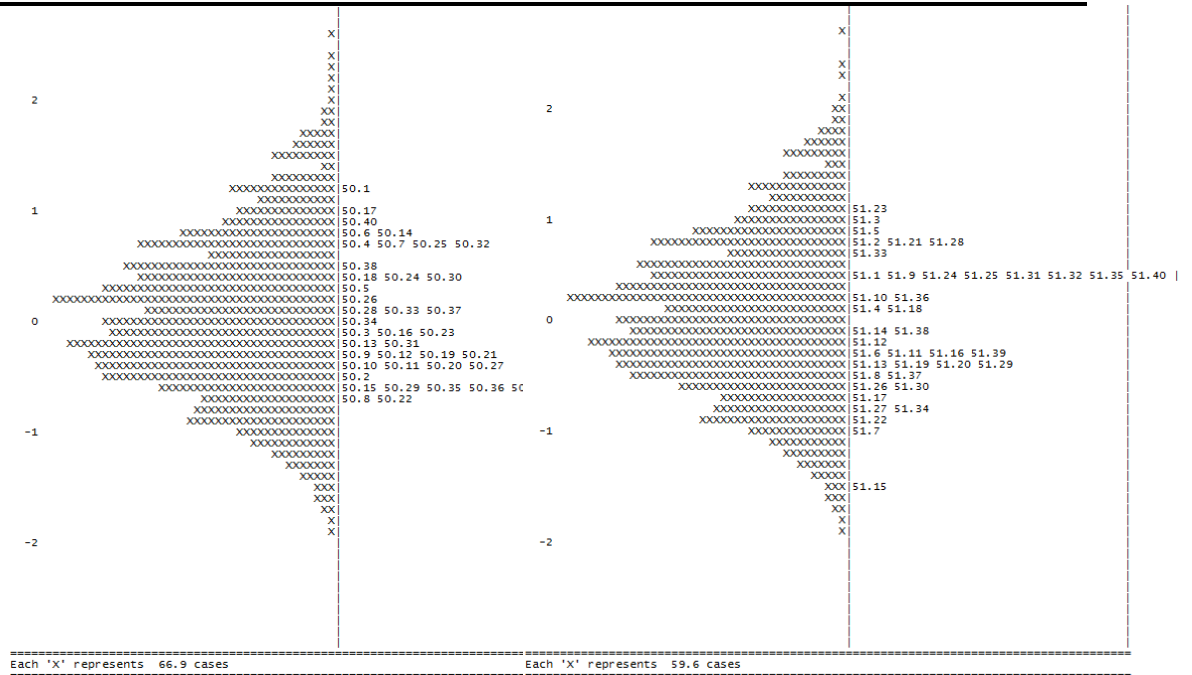
(a) ปีการศึกษา 2550

(b) ปีการศึกษา 2551



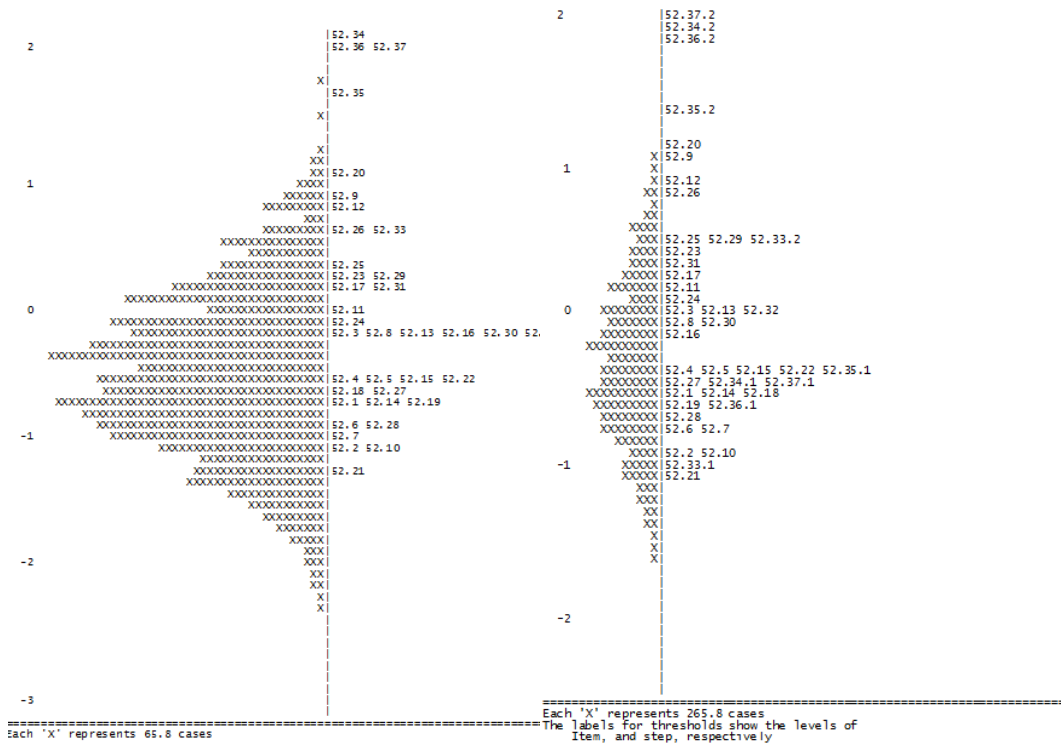
(c) ปีการศึกษา 2552

A5.9 แบบทดสอบทั้งฉบับ ปีการศึกษา 2550, 2551 และ 2552



(a) ปีการศึกษา 2550

(b) ปีการศึกษา 2551



(c) ปีการศึกษา 2552 (ให้คะแนนแบบสองค่า)

(d) ปีการศึกษา 2552 (ให้คะแนนแบบหลายค่า)

ตารางที่ B1 ค่าประมาณพารามิเตอร์ความยากข้อสอบและดัชนีความสอดคล้องกลมกลืนของ
ข้อสอบ ONET ปีการศึกษา 2550

B1.1 โมเดลเอกมิติ

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	1.120	0.005	1.15	(0.99, 1.01)	29.6	1.08	(0.99, 1.01)	20.4
2	item2	-0.559	0.005	1.01	(0.99, 1.01)	2.6	1.02	(0.99, 1.01)	7.7
3	item3	-0.190	0.005	1.01	(0.99, 1.01)	2.5	1.01	(1.00, 1.00)	4.4
4	item4	0.623	0.005	1.11	(0.99, 1.01)	21.7	1.07	(0.99, 1.01)	25.8
5	item5	0.256	0.005	0.99	(0.99, 1.01)	-2.4	0.99	(1.00, 1.00)	-5.4
6	item6	0.719	0.005	1.00	(0.99, 1.01)	-0.9	0.99	(0.99, 1.01)	-3.5
7	item7	0.695	0.005	1.05	(0.99, 1.01)	10.1	1.03	(0.99, 1.01)	9.5
8	item8	-0.747	0.005	0.93	(0.99, 1.01)	-14.9	0.95	(0.99, 1.01)	-18.2
9	item9	-0.361	0.005	0.95	(0.99, 1.01)	-10.3	0.96	(1.00, 1.00)	-16.7
10	item10	-0.481	0.005	0.93	(0.99, 1.01)	-14.5	0.95	(1.00, 1.00)	-19.4
11	item11	-0.389	0.005	0.93	(0.99, 1.01)	-15.0	0.95	(1.00, 1.00)	-23.5
12	item12	-0.359	0.005	0.93	(0.99, 1.01)	-15.2	0.95	(1.00, 1.00)	-22.7
13	item13	-0.203	0.005	0.96	(0.99, 1.01)	-8.3	0.97	(1.00, 1.00)	-12.9
14	item14	0.768	0.005	1.13	(0.99, 1.01)	25.7	1.08	(0.99, 1.01)	26.5
15	item15	-0.684	0.005	0.91	(0.99, 1.01)	-19.7	0.94	(0.99, 1.01)	-23.8
16	item16	-0.120	0.005	1.00	(0.99, 1.01)	-0.5	1.00	(1.00, 1.00)	-0.2
17	item17	0.937	0.005	1.14	(0.99, 1.01)	28.8	1.08	(0.99, 1.01)	22.8
18	item18	0.338	0.005	1.03	(0.99, 1.01)	6.0	1.02	(1.00, 1.00)	7.2
19	item19	-0.330	0.005	1.02	(0.99, 1.01)	3.4	1.01	(1.00, 1.00)	6.1
20	item20	-0.405	0.005	0.94	(0.99, 1.01)	-13.4	0.95	(1.00, 1.00)	-20.6
21	item21	-0.309	0.005	0.93	(0.99, 1.01)	-14.8	0.95	(1.00, 1.00)	-22.9
22	item22	-0.705	0.005	0.99	(0.99, 1.01)	-1.3	0.99	(0.99, 1.01)	-3.3
23	item23	-0.170	0.005	1.03	(0.99, 1.01)	5.3	1.02	(1.00, 1.00)	10.0
24	item24	0.350	0.005	0.98	(0.99, 1.01)	-4.6	0.97	(1.00, 1.00)	-10.7
25	item25	0.673	0.005	1.07	(0.99, 1.01)	14.8	1.04	(0.99, 1.01)	13.9
26	item26	0.163	0.005	1.02	(0.99, 1.01)	3.6	1.01	(1.00, 1.00)	5.3
27	item27	-0.397	0.005	0.99	(0.99, 1.01)	-2.8	0.99	(1.00, 1.00)	-5.1
28	item28	0.092	0.005	1.11	(0.99, 1.01)	22.1	1.09	(1.00, 1.00)	39.0
29	item29	-0.674	0.005	0.91	(0.99, 1.01)	-18.6	0.95	(0.99, 1.01)	-18.9
30	item30	0.411	0.005	1.05	(0.99, 1.01)	10.1	1.03	(1.00, 1.00)	12.7
31	item31	-0.174	0.005	1.01	(0.99, 1.01)	1.3	1.00	(1.00, 1.00)	-0.4
32	item32	0.618	0.005	1.08	(0.99, 1.01)	16.1	1.05	(0.99, 1.01)	18.4
33	item33	0.040	0.005	1.01	(0.99, 1.01)	1.8	1.01	(1.00, 1.00)	2.8
34	item34	-0.022	0.005	0.96	(0.99, 1.01)	-8.8	0.96	(1.00, 1.00)	-17.5
35	item35	-0.632	0.005	0.95	(0.99, 1.01)	-11.8	0.96	(0.99, 1.01)	-14.3
36	item36	-0.599	0.005	0.91	(0.99, 1.01)	-20.0	0.94	(0.99, 1.01)	-25.3
37	item37	0.056	0.005	0.98	(0.99, 1.01)	-4.5	0.98	(1.00, 1.00)	-8.6
38	item38	0.443	0.005	0.98	(0.99, 1.01)	-5.1	0.97	(1.00, 1.00)	-12.8
39	item39	-0.660	0.005	0.93	(0.99, 1.01)	-14.7	0.96	(0.99, 1.01)	-15.8
40	item40	0.869*	0.033	1.14	(0.99, 1.01)	28.3	1.08	(0.99, 1.01)	25.6

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 371920.65, df = 39, Sig Level = 0.000

[^] Quick standard errors have been used

B1.2 โมเดลพหุมิติระหว่างข้อสอบ

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	0.749	0.005	1.09	(0.99, 1.01)	19.5	1.06	(0.99, 1.01)	14.5
2	item2	-0.938	0.005	0.99	(0.99, 1.01)	-2.5	1.00	(0.99, 1.01)	-0.7
3	item3	-0.570	0.005	0.99	(0.99, 1.01)	-2.3	0.99	(1.00, 1.00)	-4.0
4	item4	0.242	0.005	1.06	(0.99, 1.01)	13.1	1.05	(0.99, 1.01)	16.4
5	item5	-0.127	0.005	0.97	(0.99, 1.01)	-5.9	0.98	(1.00, 1.00)	-10.5
6	item6	0.335	0.005	0.97	(0.99, 1.01)	-6.1	0.97	(0.99, 1.01)	-9.0
7	item7	0.310*	0.013	1.01	(0.99, 1.01)	2.9	1.01	(0.99, 1.01)	2.4
8	item8	-0.223	0.006	1.02	(0.99, 1.01)	4.2	1.01	(0.99, 1.01)	2.3
9	item9	0.223*	0.006	1.04	(0.99, 1.01)	8.0	1.03	(0.99, 1.01)	8.1
10	item10	-0.453	0.005	0.95	(0.99, 1.01)	-10.7	0.97	(0.99, 1.01)	-11.8
11	item11	-0.359	0.005	0.95	(0.99, 1.01)	-11.8	0.96	(0.99, 1.01)	-15.6
12	item12	-0.328	0.005	0.94	(0.99, 1.01)	-13.2	0.96	(0.99, 1.01)	-16.4
13	item13	-0.166	0.005	0.97	(0.99, 1.01)	-5.4	0.98	(1.00, 1.00)	-6.6
14	item14	0.833	0.005	1.15	(0.99, 1.01)	30.0	1.09	(0.99, 1.01)	28.5
15	item15	-0.648	0.005	0.92	(0.99, 1.01)	-17.7	0.94	(0.99, 1.01)	-19.2
16	item16	-0.063	0.005	1.01	(0.99, 1.01)	2.9	1.01	(1.00, 1.00)	5.4
17	item17	1.028	0.005	1.18	(0.99, 1.01)	35.9	1.09	(0.99, 1.01)	26.6
18	item18	0.419	0.005	1.05	(0.99, 1.01)	9.8	1.03	(0.99, 1.01)	11.7
19	item19	-0.262*	0.015	1.03	(0.99, 1.01)	6.3	1.03	(0.99, 1.01)	9.9
20	item20	0.069	0.005	0.99	(0.99, 1.01)	-1.2	1.00	(0.99, 1.01)	0.0
21	item21	0.186	0.005	0.98	(0.99, 1.01)	-4.3	0.99	(0.99, 1.01)	-3.2

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
22 item22	-0.255*	0.008	1.05 (0.99, 1.01)		11.4	1.03 (0.99, 1.01)		8.4
23 item23	-0.178	0.005	1.02 (0.99, 1.01)		4.0	1.02 (1.00, 1.00)		6.7
24 item24	0.353	0.005	0.99 (0.99, 1.01)		-3.1	0.98 (1.00, 1.00)		-7.5
25 item25	0.683	0.005	1.08 (0.99, 1.01)		16.5	1.05 (0.99, 1.01)		15.2
26 item26	0.160	0.005	1.02 (0.99, 1.01)		3.4	1.01 (1.00, 1.00)		4.8
27 item27	-0.410	0.005	0.98 (0.99, 1.01)		-4.9	0.98 (1.00, 1.00)		-8.4
28 item28	0.085	0.005	1.10 (0.99, 1.01)		20.9	1.08 (1.00, 1.00)		33.9
29 item29	-0.693*	0.013	0.91 (0.99, 1.01)		-19.2	0.95 (0.99, 1.01)		-19.9
30 item30	0.388	0.005	1.06 (0.99, 1.01)		12.1	1.04 (0.99, 1.01)		14.7
31 item31	-0.218	0.005	1.01 (0.99, 1.01)		1.4	1.00 (1.00, 1.00)		-0.3
32 item32	0.599	0.005	1.09 (0.99, 1.01)		18.4	1.06 (0.99, 1.01)		20.1
33 item33	-0.001	0.005	1.02 (0.99, 1.01)		3.7	1.01 (1.00, 1.00)		5.8
34 item34	-0.068	0.005	0.97 (0.99, 1.01)		-6.0	0.98 (1.00, 1.00)		-10.2
35 item35	-0.699*	0.012	0.94 (0.99, 1.01)		-12.0	0.96 (0.99, 1.01)		-13.5
36 item36	-0.657	0.005	0.93 (0.99, 1.01)		-15.8	0.95 (0.99, 1.01)		-15.7
37 item37	0.041	0.005	1.01 (0.99, 1.01)		2.2	1.01 (0.99, 1.01)		2.8
38 item38	0.453	0.005	1.02 (0.99, 1.01)		3.7	1.00 (0.99, 1.01)		-0.7
39 item39	-0.742	0.005	0.96 (0.99, 1.01)		-8.1	0.99 (0.99, 1.01)		-4.5
40 item40	0.905*	0.011	1.23 (0.99, 1.01)		46.1	1.13 (0.99, 1.01)		35.5

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 267834.81, df = 33, Sig Level = 0.000

[^] Quick standard errors have been used

B1.3 โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	0.553	0.004	1.09 (0.99, 1.01)		18.6	1.05 (0.99, 1.01)		12.9
2 item2	-0.291	0.004	0.99 (0.99, 1.01)		-2.1	1.00 (0.99, 1.01)		0.3
3 item3	-0.106	0.004	0.99 (0.99, 1.01)		-2.5	0.99 (1.00, 1.00)		-4.6
4 item4	0.301	0.004	1.06 (0.99, 1.01)		12.3	1.04 (0.99, 1.01)		14.4
5 item5	0.116	0.004	0.97 (0.99, 1.01)		-6.1	0.98 (1.00, 1.00)		-10.9
6 item6	0.347	0.004	0.97 (0.99, 1.01)		-6.7	0.97 (0.99, 1.01)		-9.9
7 item7	0.334	0.004	1.01 (0.99, 1.01)		2.6	1.00 (0.99, 1.01)		1.5
8 item8	-0.450	0.004	1.01 (0.99, 1.01)		2.4	1.01 (0.99, 1.01)		3.3
9 item9	-0.226	0.004	1.02 (0.99, 1.01)		4.5	1.02 (0.99, 1.01)		5.5
10 item10	-0.265	0.004	0.95 (0.99, 1.01)		-9.9	0.97 (0.99, 1.01)		-10.3
11 item11	-0.217	0.004	0.95 (0.99, 1.01)		-11.7	0.96 (0.99, 1.01)		-15.2
12 item12	-0.202	0.004	0.94 (0.99, 1.01)		-12.9	0.96 (0.99, 1.01)		-15.5
13 item13	-0.121	0.004	0.98 (0.99, 1.01)		-5.3	0.98 (1.00, 1.00)		-6.3
14 item14	0.377	0.004	1.15 (0.99, 1.01)		29.6	1.08 (0.99, 1.01)		26.4
15 item15	-0.363	0.004	0.92 (0.99, 1.01)		-17.5	0.94 (0.99, 1.01)		-18.4
16 item16	-0.072	0.004	1.01 (0.99, 1.01)		3.2	1.01 (1.00, 1.00)		5.8
17 item17	0.474	0.004	1.18 (0.99, 1.01)		36.1	1.08 (0.99, 1.01)		24.6
18 item18	0.169	0.004	1.05 (0.99, 1.01)		10.2	1.03 (1.00, 1.00)		10.9
19 item19	-0.170	0.004	1.03 (0.99, 1.01)		6.1	1.03 (0.99, 1.01)		10.0
20 item20	-0.230	0.004	1.02 (0.99, 1.01)		5.2	1.02 (0.99, 1.01)		4.9
21 item21	-0.171	0.004	1.01 (0.99, 1.01)		1.7	1.01 (0.99, 1.01)		1.9
22 item22	-0.393	0.004	1.10 (0.99, 1.01)		19.8	1.05 (0.99, 1.01)		14.2
23 item23	-0.072	0.004	1.02 (0.99, 1.01)		3.6	1.01 (1.00, 1.00)		6.1
24 item24	0.193	0.004	0.99 (0.99, 1.01)		-3.0	0.98 (1.00, 1.00)		-8.0
25 item25	0.358	0.004	1.08 (0.99, 1.01)		15.9	1.04 (0.99, 1.01)		13.5
26 item26	0.097	0.004	1.01 (0.99, 1.01)		2.9	1.01 (1.00, 1.00)		3.6
27 item27	-0.187	0.004	0.97 (0.99, 1.01)		-5.6	0.98 (1.00, 1.00)		-8.7
28 item28	0.060	0.004	1.09 (0.99, 1.01)		19.1	1.08 (1.00, 1.00)		31.8
29 item29	-0.328	0.004	0.91 (0.99, 1.01)		-18.6	0.95 (0.99, 1.01)		-19.8
30 item30	0.222	0.004	1.06 (0.99, 1.01)		13.4	1.04 (0.99, 1.01)		14.9
31 item31	-0.081	0.004	1.01 (0.99, 1.01)		3.1	1.00 (1.00, 1.00)		0.8
32 item32	0.327	0.004	1.09 (0.99, 1.01)		19.4	1.06 (0.99, 1.01)		19.7
33 item33	0.027	0.004	1.02 (0.99, 1.01)		5.1	1.02 (1.00, 1.00)		7.8
34 item34	-0.006	0.004	0.98 (0.99, 1.01)		-4.5	0.98 (1.00, 1.00)		-8.3
35 item35	-0.322	0.004	0.95 (0.99, 1.01)		-10.9	0.97 (0.99, 1.01)		-12.0
36 item36	-0.323	0.004	0.93 (0.99, 1.01)		-15.1	0.96 (0.99, 1.01)		-14.0
37 item37	0.028	0.004	1.01 (0.99, 1.01)		1.6	1.00 (0.99, 1.01)		1.5
38 item38	0.233	0.004	1.02 (0.99, 1.01)		3.5	1.00 (0.99, 1.01)		-1.3
39 item39	-0.365	0.004	0.97 (0.99, 1.01)		-6.7	0.99 (0.99, 1.01)		-2.5
40 item40	0.458	0.004	1.24 (0.99, 1.01)		46.5	1.12 (0.99, 1.01)		33.9

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 212744.50, df = 40, Sig Level = 0.000

[^] Quick standard errors have been used

B1.4 โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	1.116	0.008	1.10	(0.99, 1.01)	21.3	1.05	(0.99, 1.01)	12.8
2 item2	-0.584	0.007	1.01	(0.99, 1.01)	1.7	1.02	(0.99, 1.01)	5.6
3 item3	-0.212	0.007	1.00	(0.99, 1.01)	0.1	1.00	(1.00, 1.00)	-0.9
4 item4	0.608	0.007	1.07	(0.99, 1.01)	14.4	1.04	(0.99, 1.01)	14.8
5 item5	0.235	0.007	0.98	(0.99, 1.01)	-4.6	0.98	(1.00, 1.00)	-8.5
6 item6	0.701	0.007	0.98	(0.99, 1.01)	-4.9	0.98	(0.99, 1.01)	-8.8
7 item7	0.675	0.007	1.02	(0.99, 1.01)	4.7	1.01	(0.99, 1.01)	2.3
8 item8	-0.455	0.004	1.01	(0.99, 1.01)	3.0	1.01	(0.99, 1.01)	3.1
9 item9	-0.440	0.008	1.01	(0.99, 1.01)	1.5	1.01	(0.99, 1.01)	2.1
10 item10	-0.276	0.004	0.97	(0.99, 1.01)	-7.1	0.98	(0.99, 1.01)	-5.9
11 item11	-0.429	0.007	0.95	(0.99, 1.01)	-11.6	0.96	(0.99, 1.01)	-15.4
12 item12	-0.398	0.007	0.94	(0.99, 1.01)	-13.3	0.96	(0.99, 1.01)	-16.7
13 item13	-0.237	0.007	0.97	(0.99, 1.01)	-6.3	0.98	(1.00, 1.00)	-8.1
14 item14	0.760	0.008	1.14	(0.99, 1.01)	28.2	1.08	(0.99, 1.01)	25.1
15 item15	-0.720	0.007	0.92	(0.99, 1.01)	-17.2	0.95	(0.99, 1.01)	-17.6
16 item16	-0.137	0.007	1.01	(0.99, 1.01)	2.7	1.01	(1.00, 1.00)	4.3
17 item17	0.953	0.008	1.17	(0.99, 1.01)	34.3	1.08	(0.99, 1.01)	23.3
18 item18	0.345	0.007	1.05	(0.99, 1.01)	9.7	1.02	(1.00, 1.00)	9.5
19 item19	-0.335	0.007	1.02	(0.99, 1.01)	5.2	1.02	(1.00, 1.00)	8.0
20 item20	-0.447	0.008	1.03	(0.99, 1.01)	5.6	1.02	(0.99, 1.01)	6.4
21 item21	-0.330	0.008	1.01	(0.99, 1.01)	1.3	1.01	(0.99, 1.01)	3.2
22 item22	-0.772	0.008	1.10	(0.99, 1.01)	20.6	1.05	(0.99, 1.01)	14.5
23 item23	-0.138	0.007	1.01	(0.99, 1.01)	2.7	1.01	(1.00, 1.00)	4.3
24 item24	0.393	0.007	0.98	(0.99, 1.01)	-3.4	0.98	(1.00, 1.00)	-10.0
25 item25	0.722	0.007	1.08	(0.99, 1.01)	15.6	1.03	(0.99, 1.01)	11.8
26 item26	0.106	0.004	1.03	(0.99, 1.01)	6.6	1.02	(0.99, 1.01)	6.8
27 item27	-0.370	0.007	0.97	(0.99, 1.01)	-5.8	0.98	(1.00, 1.00)	-9.1
28 item28	0.125	0.007	1.09	(0.99, 1.01)	18.3	1.07	(1.00, 1.00)	29.6
29 item29	-0.652	0.007	0.92	(0.99, 1.01)	-17.1	0.95	(0.99, 1.01)	-16.6
30 item30	0.453	0.007	1.06	(0.99, 1.01)	12.5	1.03	(0.99, 1.01)	12.3
31 item31	-0.077	0.004	1.02	(0.99, 1.01)	4.0	1.01	(0.99, 1.01)	4.0
32 item32	0.663	0.007	1.09	(0.99, 1.01)	17.9	1.05	(0.99, 1.01)	16.2
33 item33	0.035	0.004	1.05	(0.99, 1.01)	9.4	1.03	(0.99, 1.01)	10.5
34 item34	-0.001	0.007	0.98	(0.99, 1.01)	-4.7	0.98	(1.00, 1.00)	-9.2
35 item35	-0.631	0.007	0.95	(0.99, 1.01)	-11.4	0.96	(0.99, 1.01)	-12.6
36 item36	-0.313	0.004	0.93	(0.99, 1.01)	-14.1	0.96	(0.99, 1.01)	-12.6
37 item37	0.033	0.004	1.00	(0.99, 1.01)	1.0	1.00	(0.99, 1.01)	-0.3
38 item38	0.235	0.004	1.02	(0.99, 1.01)	3.6	0.99	(0.99, 1.01)	-3.4
39 item39	-0.355	0.004	0.97	(0.99, 1.01)	-5.6	1.00	(0.99, 1.01)	-1.5
40 item40	0.458	0.004	1.22	(0.99, 1.01)	43.3	1.10	(0.99, 1.01)	29.3

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 212424.24, df = 40, Sig Level = 0.000

[^] Quick standard errors have been used

ตารางที่ B2 ค่าประมาณพารามิเตอร์ความยากข้อสอบและดัชนีความสอดคล้องกลมกลืนของ
ข้อสอบ ONET ปีการศึกษา 2551

B2.1 โมเดลเอกมิติ

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	0.511	0.005	1.13	(0.99, 1.01)	25.5	1.10	(0.99, 1.01)	38.0
2	item2	0.716	0.006	0.99	(0.99, 1.01)	-1.4	0.97	(0.99, 1.01)	-9.1
3	item3	0.930	0.006	1.06	(0.99, 1.01)	11.4	1.02	(0.99, 1.01)	4.5
4	item4	0.189	0.005	1.04	(0.99, 1.01)	7.3	1.03	(1.00, 1.00)	12.4
5	item5	0.827	0.006	1.07	(0.99, 1.01)	14.3	1.04	(0.99, 1.01)	13.8
6	item6	-0.230	0.005	0.96	(0.99, 1.01)	-7.6	0.97	(1.00, 1.00)	-10.9
7	item7	-0.982	0.006	0.84	(0.99, 1.01)	-32.4	0.91	(0.99, 1.01)	-24.4
8	item8	-0.467	0.006	0.95	(0.99, 1.01)	-9.1	0.97	(0.99, 1.01)	-11.7
9	item9	0.474	0.005	0.99	(0.99, 1.01)	-2.5	0.98	(0.99, 1.01)	-8.8
10	item10	0.192	0.005	1.05	(0.99, 1.01)	10.5	1.05	(1.00, 1.00)	19.0
11	item11	-0.196	0.005	1.02	(0.99, 1.01)	3.3	1.01	(1.00, 1.00)	6.0
12	item12	-0.156	0.005	0.97	(0.99, 1.01)	-6.8	0.98	(1.00, 1.00)	-9.8
13	item13	-0.382	0.005	0.90	(0.99, 1.01)	-20.1	0.92	(0.99, 1.01)	-29.7
14	item14	-0.012	0.005	1.00	(0.99, 1.01)	0.7	1.00	(1.00, 1.00)	1.5
15	item15	-1.452	0.006	0.74	(0.99, 1.01)	-55.8	0.86	(0.99, 1.01)	-27.7
16	item16	-0.260	0.005	1.02	(0.99, 1.01)	3.8	1.01	(1.00, 1.00)	5.5
17	item17	-0.679	0.006	0.91	(0.99, 1.01)	-17.5	0.94	(0.99, 1.01)	-19.3
18	item18	0.118	0.005	0.99	(0.99, 1.01)	-2.5	0.99	(1.00, 1.00)	-5.9
19	item19	-0.336	0.005	0.99	(0.99, 1.01)	-2.7	0.99	(0.99, 1.01)	-2.4
20	item20	-0.363	0.005	0.97	(0.99, 1.01)	-5.6	0.98	(0.99, 1.01)	-8.5
21	item21	0.733	0.006	1.02	(0.99, 1.01)	4.1	0.99	(0.99, 1.01)	-2.9
22	item22	-0.873	0.006	0.90	(0.99, 1.01)	-21.3	0.94	(0.99, 1.01)	-18.2
23	item23	1.112	0.006	1.11	(0.99, 1.01)	20.9	1.06	(0.99, 1.01)	14.9
24	item24	0.490	0.005	1.01	(0.99, 1.01)	2.9	1.00	(0.99, 1.01)	1.0
25	item25	0.475	0.005	1.06	(0.99, 1.01)	11.2	1.04	(0.99, 1.01)	16.0
26	item26	-0.599	0.006	0.93	(0.99, 1.01)	-13.6	0.96	(0.99, 1.01)	-14.0
27	item27	-0.697	0.006	1.04	(0.99, 1.01)	7.5	1.02	(0.99, 1.01)	5.7
28	item28	0.714	0.006	1.01	(0.99, 1.01)	2.0	0.99	(0.99, 1.01)	-4.8
29	item29	-0.318	0.005	1.01	(0.99, 1.01)	2.8	1.00	(0.99, 1.01)	0.7
30	item30	-0.529	0.006	0.84	(0.99, 1.01)	-32.6	0.89	(0.99, 1.01)	-39.7
31	item31	0.464	0.005	1.07	(0.99, 1.01)	13.3	1.05	(0.99, 1.01)	19.8
32	item32	0.470	0.005	1.07	(0.99, 1.01)	13.4	1.05	(0.99, 1.01)	19.3
33	item33	0.601	0.005	1.05	(0.99, 1.01)	10.0	1.04	(0.99, 1.01)	12.8
34	item34	-0.779	0.006	0.92	(0.99, 1.01)	-16.0	0.96	(0.99, 1.01)	-13.2
35	item35	0.401	0.005	1.09	(0.99, 1.01)	17.9	1.08	(1.00, 1.00)	30.6
36	item36	0.220	0.005	1.01	(0.99, 1.01)	2.8	1.01	(1.00, 1.00)	4.4
37	item37	-0.470	0.005	0.98	(0.99, 1.01)	-4.8	0.99	(0.99, 1.01)	-4.9
38	item38	-0.063	0.005	0.93	(0.99, 1.01)	-15.0	0.94	(1.00, 1.00)	-25.3
39	item39	-0.307	0.005	0.98	(0.99, 1.01)	-3.5	0.98	(0.99, 1.01)	-6.2
40	item40	0.512*	0.034	1.14	(0.99, 1.01)	26.7	1.11	(0.99, 1.01)	41.0

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 430714.10, df = 39, Sig Level = 0.000

[^] Quick standard errors have been used

B2.2 โมเดลพหุมิติระหว่างข้อสอบ

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	0.246	0.006	1.14	(0.99, 1.01)	25.7	1.10	(0.99, 1.01)	33.9
2	item2	0.456	0.006	1.00	(0.99, 1.01)	-0.1	0.98	(0.99, 1.01)	-5.3
3	item3	0.674	0.006	1.07	(0.99, 1.01)	13.3	1.03	(0.99, 1.01)	7.3
4	item4	-0.093	0.006	1.03	(0.99, 1.01)	6.4	1.02	(1.00, 1.00)	9.8
5	item5	0.556	0.006	1.07	(0.99, 1.01)	14.3	1.04	(0.99, 1.01)	13.1
6	item6	-0.532	0.006	0.96	(0.99, 1.01)	-8.0	0.97	(0.99, 1.01)	-11.0
7	item7	-1.306*	0.014	0.85	(0.99, 1.01)	-30.3	0.91	(0.99, 1.01)	-22.8
8	item8	-0.568	0.006	0.97	(0.99, 1.01)	-5.7	0.98	(0.99, 1.01)	-6.7
9	item9	0.439	0.006	1.01	(0.99, 1.01)	1.2	1.00	(0.99, 1.01)	-0.4
10	item10	0.128*	0.008	1.08	(0.99, 1.01)	16.0	1.06	(0.99, 1.01)	20.8
11	item11	0.219	0.006	1.08	(0.99, 1.01)	14.5	1.05	(0.99, 1.01)	18.6
12	item12	0.252	0.006	1.01	(0.99, 1.01)	2.1	1.01	(0.99, 1.01)	4.8
13	item13	0.001	0.006	0.93	(0.99, 1.01)	-14.9	0.94	(0.99, 1.01)	-18.2
14	item14	0.392	0.006	1.06	(0.99, 1.01)	10.8	1.04	(0.99, 1.01)	14.0
15	item15	-1.176	0.006	0.77	(0.99, 1.01)	-48.4	0.90	(0.99, 1.01)	-17.6
16	item16	0.118	0.006	1.08	(0.99, 1.01)	15.6	1.05	(0.99, 1.01)	16.3
17	item17	-0.333	0.006	0.96	(0.99, 1.01)	-7.6	0.98	(0.99, 1.01)	-6.8
18	item18	0.526*	0.015	1.03	(0.99, 1.01)	6.6	1.02	(0.99, 1.01)	5.6
19	item19	-0.382	0.006	1.04	(0.99, 1.01)	8.4	1.04	(0.99, 1.01)	11.3
20	item20	-0.411	0.006	1.02	(0.99, 1.01)	3.8	1.01	(0.99, 1.01)	4.3
21	item21	0.793*	0.008	1.05	(0.99, 1.01)	10.7	1.02	(0.99, 1.01)	5.1

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
22 item22	-0.903	0.006	0.91 (0.99, 1.01)	-18.7	0.95 (0.99, 1.01)	-14.3		
23 item23	1.152	0.006	1.11 (0.99, 1.01)	20.7	1.05 (0.99, 1.01)	13.3		
24 item24	0.521	0.006	1.02 (0.99, 1.01)	3.1	1.01 (0.99, 1.01)	2.4		
25 item25	0.511	0.006	1.06 (0.99, 1.01)	12.3	1.04 (0.99, 1.01)	15.2		
26 item26	-0.593	0.006	0.94 (0.99, 1.01)	-12.7	0.96 (0.99, 1.01)	-12.4		
27 item27	-0.688*	0.013	1.04 (0.99, 1.01)	7.8	1.02 (0.99, 1.01)	5.0		
28 item28	0.477	0.006	1.02 (0.99, 1.01)	3.7	1.00 (0.99, 1.01)	-0.8		
29 item29	-0.581	0.006	1.03 (0.99, 1.01)	5.0	1.01 (0.99, 1.01)	2.3		
30 item30	-0.793	0.006	0.85 (0.99, 1.01)	-31.4	0.90 (0.99, 1.01)	-35.3		
31 item31	0.240	0.006	1.08 (0.99, 1.01)	16.0	1.06 (0.99, 1.01)	20.6		
32 item32	0.255	0.006	1.09 (0.99, 1.01)	17.3	1.06 (0.99, 1.01)	22.0		
33 item33	0.400*	0.013	1.06 (0.99, 1.01)	12.5	1.04 (0.99, 1.01)	14.2		
34 item34	-0.726	0.006	0.91 (0.99, 1.01)	-19.4	0.94 (0.99, 1.01)	-18.4		
35 item35	0.489	0.006	1.10 (0.99, 1.01)	19.7	1.08 (0.99, 1.01)	28.9		
36 item36	0.302	0.006	1.02 (0.99, 1.01)	4.4	1.02 (0.99, 1.01)	6.4		
37 item37	-0.412	0.006	0.96 (0.99, 1.01)	-7.8	0.97 (0.99, 1.01)	-9.0		
38 item38	0.003	0.006	0.93 (0.99, 1.01)	-14.1	0.94 (0.99, 1.01)	-21.5		
39 item39	-0.250	0.006	0.98 (0.99, 1.01)	-3.5	0.99 (0.99, 1.01)	-5.3		
40 item40	0.594*	0.014	1.17 (0.99, 1.01)	31.8	1.12 (0.99, 1.01)	41.5		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 279502.98, df = 33, Sig Level = 0.000

[^] Quick standard errors have been used

B2.3 โมเดลพหุมิติภายในข้อสอบโดยข้อสอบทุกข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	0.222	0.004	1.14 (0.99, 1.01)	25.9	1.10 (0.99, 1.01)	34.6		
2 item2	0.327	0.004	0.99 (0.99, 1.01)	-1.4	0.98 (0.99, 1.01)	-5.8		
3 item3	0.436	0.004	1.06 (0.99, 1.01)	11.8	1.02 (0.99, 1.01)	6.9		
4 item4	0.052	0.004	1.03 (0.99, 1.01)	6.4	1.03 (1.00, 1.00)	10.5		
5 item5	0.376	0.004	1.07 (0.99, 1.01)	14.0	1.04 (0.99, 1.01)	13.1		
6 item6	-0.168	0.004	0.96 (0.99, 1.01)	-7.7	0.97 (0.99, 1.01)	-10.0		
7 item7	-0.555	0.004	0.86 (0.99, 1.01)	-29.8	0.92 (0.99, 1.01)	-22.0		
8 item8	-0.314	0.004	0.98 (0.99, 1.01)	-3.2	0.99 (0.99, 1.01)	-3.2		
9 item9	0.192	0.004	1.01 (0.99, 1.01)	1.8	1.00 (0.99, 1.01)	0.7		
10 item10	0.035	0.004	1.10 (0.99, 1.01)	18.8	1.07 (0.99, 1.01)	24.1		
11 item11	-0.175	0.004	1.08 (0.99, 1.01)	15.8	1.06 (0.99, 1.01)	20.3		
12 item12	-0.158	0.004	1.01 (0.99, 1.01)	2.6	1.02 (0.99, 1.01)	5.7		
13 item13	-0.283	0.004	0.93 (0.99, 1.01)	-15.2	0.94 (0.99, 1.01)	-18.6		
14 item14	-0.088	0.004	1.06 (0.99, 1.01)	11.6	1.05 (0.99, 1.01)	15.5		
15 item15	-0.869	0.005	0.77 (0.99, 1.01)	-50.6	0.90 (0.99, 1.01)	-19.5		
16 item16	-0.225	0.004	1.09 (0.99, 1.01)	16.6	1.06 (0.99, 1.01)	18.0		
17 item17	-0.450	0.004	0.96 (0.99, 1.01)	-7.8	0.97 (0.99, 1.01)	-7.6		
18 item18	-0.021	0.004	1.04 (0.99, 1.01)	8.3	1.02 (0.99, 1.01)	7.8		
19 item19	-0.265	0.004	1.03 (0.99, 1.01)	5.2	1.03 (0.99, 1.01)	7.8		
20 item20	-0.279	0.004	1.00 (0.99, 1.01)	0.6	1.00 (0.99, 1.01)	0.1		
21 item21	0.318	0.004	1.05 (0.99, 1.01)	9.1	1.01 (0.99, 1.01)	3.0		
22 item22	-0.527	0.004	0.91 (0.99, 1.01)	-18.3	0.95 (0.99, 1.01)	-13.7		
23 item23	0.500	0.004	1.11 (0.99, 1.01)	20.6	1.05 (0.99, 1.01)	13.0		
24 item24	0.184	0.004	1.02 (0.99, 1.01)	3.4	1.01 (0.99, 1.01)	2.4		
25 item25	0.178	0.004	1.06 (0.99, 1.01)	12.2	1.04 (0.99, 1.01)	15.2		
26 item26	-0.374	0.004	0.94 (0.99, 1.01)	-12.2	0.96 (0.99, 1.01)	-11.3		
27 item27	-0.421	0.004	1.05 (0.99, 1.01)	9.2	1.02 (0.99, 1.01)	6.5		
28 item28	0.309	0.004	1.01 (0.99, 1.01)	1.0	0.99 (0.99, 1.01)	-3.5		
29 item29	-0.215	0.004	1.01 (0.99, 1.01)	2.4	1.00 (0.99, 1.01)	0.0		
30 item30	-0.319	0.004	0.84 (0.99, 1.01)	-32.6	0.89 (0.99, 1.01)	-38.2		
31 item31	0.192	0.004	1.07 (0.99, 1.01)	13.4	1.05 (0.99, 1.01)	18.5		
32 item32	0.200	0.004	1.08 (0.99, 1.01)	14.8	1.06 (0.99, 1.01)	20.0		
33 item33	0.271	0.004	1.05 (0.99, 1.01)	10.5	1.04 (0.99, 1.01)	12.9		
34 item34	-0.431	0.004	0.90 (0.99, 1.01)	-19.5	0.94 (0.99, 1.01)	-18.0		
35 item35	0.174	0.004	1.10 (0.99, 1.01)	19.9	1.09 (0.99, 1.01)	29.5		
36 item36	0.081	0.004	1.02 (0.99, 1.01)	3.8	1.02 (0.99, 1.01)	5.9		
37 item37	-0.274	0.004	0.96 (0.99, 1.01)	-7.6	0.97 (0.99, 1.01)	-8.8		
38 item38	-0.067	0.004	0.93 (0.99, 1.01)	-14.5	0.94 (0.99, 1.01)	-21.5		
39 item39	-0.193	0.004	0.98 (0.99, 1.01)	-3.5	0.99 (0.99, 1.01)	-5.4		
40 item40	0.227	0.004	1.17 (0.99, 1.01)	31.9	1.13 (0.99, 1.01)	42.4		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 230396.02, df = 40, Sig Level = 0.000

[^] Quick standard errors have been used

B2.4 โมเดลพหุมิติภายในข้อสอบโดยข้อสอบบางข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	0.445	0.008	1.15	(0.99, 1.01)	27.4	1.10	(0.99, 1.01)	34.9
2 item2	0.654	0.008	1.01	(0.99, 1.01)	2.6	0.99	(0.99, 1.01)	-4.4
3 item3	0.447	0.004	1.08	(0.99, 1.01)	15.4	1.03	(0.99, 1.01)	8.3
4 item4	0.106	0.008	1.05	(0.99, 1.01)	8.8	1.03	(1.00, 1.00)	12.9
5 item5	0.752	0.008	1.08	(0.99, 1.01)	15.9	1.04	(0.99, 1.01)	12.8
6 item6	-0.332	0.008	0.97	(0.99, 1.01)	-6.6	0.98	(0.99, 1.01)	-8.3
7 item7	-1.105	0.009	0.85	(0.99, 1.01)	-30.4	0.91	(0.99, 1.01)	-23.7
8 item8	-0.613	0.008	0.98	(0.99, 1.01)	-3.7	0.99	(0.99, 1.01)	-2.9
9 item9	0.382	0.008	1.01	(0.99, 1.01)	1.7	0.99	(0.99, 1.01)	-2.7
10 item10	0.074	0.008	1.08	(0.99, 1.01)	16.1	1.06	(0.99, 1.01)	21.2
11 item11	-0.348	0.008	1.08	(0.99, 1.01)	16.2	1.06	(0.99, 1.01)	21.0
12 item12	-0.315	0.008	1.03	(0.99, 1.01)	4.9	1.02	(0.99, 1.01)	7.3
13 item13	-0.566	0.008	0.93	(0.99, 1.01)	-13.2	0.95	(0.99, 1.01)	-14.1
14 item14	-0.094	0.004	1.06	(0.99, 1.01)	11.7	1.04	(0.99, 1.01)	14.3
15 item15	-1.741	0.010	0.77	(0.99, 1.01)	-48.4	0.90	(0.99, 1.01)	-18.8
16 item16	-0.449	0.008	1.08	(0.99, 1.01)	15.6	1.06	(0.99, 1.01)	18.5
17 item17	-0.901	0.008	0.97	(0.99, 1.01)	-6.0	0.99	(0.99, 1.01)	-2.9
18 item18	-0.025	0.004	1.03	(0.99, 1.01)	6.6	1.02	(0.99, 1.01)	6.9
19 item19	-0.535	0.008	1.04	(0.99, 1.01)	7.6	1.03	(0.99, 1.01)	10.3
20 item20	-0.564	0.008	1.01	(0.99, 1.01)	2.6	1.01	(0.99, 1.01)	2.5
21 item21	0.317	0.004	1.05	(0.99, 1.01)	10.4	1.01	(0.99, 1.01)	1.8
22 item22	-1.057	0.008	0.92	(0.99, 1.01)	-15.7	0.96	(0.99, 1.01)	-10.3
23 item23	1.001	0.008	1.12	(0.99, 1.01)	22.4	1.04	(0.99, 1.01)	10.8
24 item24	0.184	0.004	1.02	(0.99, 1.01)	4.8	1.01	(0.99, 1.01)	3.4
25 item25	0.358	0.008	1.08	(0.99, 1.01)	14.9	1.04	(0.99, 1.01)	15.8
26 item26	-0.748	0.008	0.95	(0.99, 1.01)	-9.7	0.98	(0.99, 1.01)	-7.1
27 item27	-0.431	0.004	1.06	(0.99, 1.01)	11.9	1.03	(0.99, 1.01)	8.5
28 item28	0.311	0.004	1.02	(0.99, 1.01)	4.7	0.99	(0.99, 1.01)	-1.7
29 item29	-0.432	0.008	1.02	(0.99, 1.01)	3.6	1.01	(0.99, 1.01)	2.0
30 item30	-0.327	0.004	0.85	(0.99, 1.01)	-30.6	0.90	(0.99, 1.01)	-34.0
31 item31	0.192	0.004	1.08	(0.99, 1.01)	14.9	1.05	(0.99, 1.01)	18.4
32 item32	0.402	0.008	1.09	(0.99, 1.01)	17.6	1.06	(0.99, 1.01)	19.8
33 item33	0.546	0.008	1.07	(0.99, 1.01)	13.3	1.04	(0.99, 1.01)	12.3
34 item34	-0.866	0.008	0.91	(0.99, 1.01)	-18.0	0.95	(0.99, 1.01)	-15.4
35 item35	0.345	0.008	1.10	(0.99, 1.01)	19.9	1.08	(0.99, 1.01)	28.4
36 item36	0.158	0.008	1.02	(0.99, 1.01)	4.1	1.01	(1.00, 1.00)	5.2
37 item37	-0.553	0.008	0.96	(0.99, 1.01)	-7.0	0.98	(0.99, 1.01)	-7.3
38 item38	-0.139	0.008	0.93	(0.99, 1.01)	-14.2	0.95	(1.00, 1.00)	-21.4
39 item39	-0.391	0.008	0.99	(0.99, 1.01)	-2.8	0.99	(0.99, 1.01)	-5.1
40 item40	0.451	0.008	1.17	(0.99, 1.01)	32.5	1.12	(0.99, 1.01)	40.8

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 231318.13, df = 40, Sig Level = 0.000

[^] Quick standard errors have been used

ตารางที่ B3 ค่าประมาณพารามิเตอร์ความยากข้อสอบและดัชนีความสอดคล้องกลมกลืนของ
ข้อสอบ ONET ปีการศึกษา 2552

B3.1D โมเดลเอกมิตีที่ให้คะแนนคำตอบแบบสองค่า

VARIABLES			UNWEIGHTED FIT				WEIGHTED FIT			
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T		
1 item1	-0.770	0.006	1.05	(0.99, 1.01)	11.1	1.05	(1.00, 1.00)	22.2		
2 item2	-1.091	0.006	0.91	(0.99, 1.01)	-20.4	0.93	(1.00, 1.00)	-28.3		
3 item3	-0.150	0.006	1.07	(0.99, 1.01)	14.2	1.06	(1.00, 1.00)	24.6		
4 item4	-0.543	0.006	0.94	(0.99, 1.01)	-12.9	0.95	(1.00, 1.00)	-26.4		
5 item5	-0.527	0.006	1.00	(0.99, 1.01)	-0.0	1.00	(1.00, 1.00)	-0.7		
6 item6	-0.951	0.006	0.92	(0.99, 1.01)	-17.6	0.94	(1.00, 1.00)	-28.5		
7 item7	-0.990	0.006	0.92	(0.99, 1.01)	-16.9	0.94	(1.00, 1.00)	-26.0		
8 item8	-0.199	0.006	1.00	(0.99, 1.01)	-0.4	1.00	(1.00, 1.00)	-1.8		
9 item9	0.862	0.007	1.03	(0.99, 1.01)	5.9	0.98	(0.99, 1.01)	-3.5		
10 item10	-1.113	0.006	0.97	(0.99, 1.01)	-7.2	0.97	(0.99, 1.01)	-10.4		
11 item11	-0.016	0.006	1.01	(0.99, 1.01)	2.4	1.00	(1.00, 1.00)	0.8		
12 item12	0.797	0.007	1.19	(0.99, 1.01)	37.9	1.10	(0.99, 1.01)	23.2		
13 item13	-0.145	0.006	1.01	(0.99, 1.01)	1.6	1.00	(1.00, 1.00)	1.3		
14 item14	-0.689	0.006	0.97	(0.99, 1.01)	-6.2	0.98	(1.00, 1.00)	-11.8		
15 item15	-0.513	0.006	0.96	(0.99, 1.01)	-9.1	0.96	(1.00, 1.00)	-19.8		
16 item16	-0.217	0.006	0.90	(0.99, 1.01)	-22.4	0.90	(1.00, 1.00)	-46.2		
17 item17	0.155	0.006	1.06	(0.99, 1.01)	11.7	1.03	(0.99, 1.01)	10.0		
18 item18	-0.649	0.006	0.97	(0.99, 1.01)	-7.2	0.97	(1.00, 1.00)	-14.6		
19 item19	-0.739	0.006	0.98	(0.99, 1.01)	-3.4	0.99	(1.00, 1.00)	-6.0		
20 item20	1.033	0.007	1.21	(0.99, 1.01)	40.7	1.07	(0.99, 1.01)	13.3		
21 item21	-1.222	0.006	0.94	(0.99, 1.01)	-13.1	0.96	(0.99, 1.01)	-16.3		
22 item22	-0.548	0.006	0.95	(0.99, 1.01)	-10.1	0.96	(1.00, 1.00)	-21.0		
23 item23	0.299	0.006	1.08	(0.99, 1.01)	16.2	1.05	(0.99, 1.01)	17.0		
24 item24	-0.084	0.006	1.01	(0.99, 1.01)	2.6	1.00	(1.00, 1.00)	1.9		
25 item25	0.335	0.006	1.10	(0.99, 1.01)	20.8	1.07	(0.99, 1.01)	21.2		
26 item26	0.636	0.007	1.13	(0.99, 1.01)	26.0	1.06	(0.99, 1.01)	15.6		
27 item27	-0.646	0.006	0.90	(0.99, 1.01)	-20.9	0.92	(1.00, 1.00)	-41.5		
28 item28	-0.879	0.006	0.92	(0.99, 1.01)	-16.2	0.94	(1.00, 1.00)	-25.4		
29 item29	0.272	0.006	1.09	(0.99, 1.01)	18.5	1.06	(0.99, 1.01)	19.1		
30 item30	-0.232	0.006	0.99	(0.99, 1.01)	-2.6	0.98	(1.00, 1.00)	-7.6		
31 item31	0.131	0.006	1.10	(0.99, 1.01)	21.0	1.08	(0.99, 1.01)	28.9		
32 item32	-0.152	0.006	1.04	(0.99, 1.01)	8.6	1.03	(1.00, 1.00)	14.5		
33 item33	0.572	0.007	0.88	(0.99, 1.01)	-26.9	0.92	(0.99, 1.01)	-23.3		
34 item34	2.222	0.009	1.11	(0.99, 1.01)	21.7	1.00	(0.98, 1.02)	0.2		
35 item35	1.646	0.008	0.90	(0.99, 1.01)	-22.0	0.96	(0.99, 1.01)	-5.7		
36 item36	2.049	0.009	1.37	(0.99, 1.01)	69.7	1.07	(0.98, 1.02)	7.1		
37 item37	2.058*	0.039	1.25	(0.99, 1.01)	49.3	1.05	(0.98, 1.02)	5.1		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 500439.14, df = 36, Sig Level = 0.000

[^] Quick standard errors have been used

B3.1P โมเดลเอกมิตีที่ให้คะแนนคำตอบแบบหลายค่า

VARIABLES			UNWEIGHTED FIT				WEIGHTED FIT			
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T		
1 item1	-0.555	0.005	1.03	(0.99, 1.01)	5.6	1.03	(1.00, 1.00)	13.1		
2 item2	-0.870	0.005	0.90	(0.99, 1.01)	-20.7	0.93	(1.00, 1.00)	-32.2		
3 item3	0.049	0.005	1.04	(0.99, 1.01)	9.3	1.04	(1.00, 1.00)	17.6		
4 item4	-0.337	0.005	0.94	(0.99, 1.01)	-13.5	0.94	(1.00, 1.00)	-32.2		
5 item5	-0.323	0.005	0.99	(0.99, 1.01)	-2.9	0.99	(1.00, 1.00)	-7.4		
6 item6	-0.739	0.005	0.92	(0.99, 1.01)	-17.3	0.93	(1.00, 1.00)	-31.4		
7 item7	-0.779	0.005	0.92	(0.99, 1.01)	-17.1	0.94	(1.00, 1.00)	-29.4		
8 item8	-0.007	0.005	0.99	(0.99, 1.01)	-3.1	0.99	(1.00, 1.00)	-7.5		
9 item9	1.019	0.005	0.99	(0.99, 1.01)	-1.6	0.97	(0.99, 1.01)	-6.5		
10 item10	-0.921	0.005	0.96	(0.99, 1.01)	-8.6	0.97	(1.00, 1.00)	-12.6		
11 item11	0.146	0.005	0.99	(0.99, 1.01)	-2.5	0.98	(1.00, 1.00)	-6.9		
12 item12	0.924	0.005	1.11	(0.99, 1.01)	22.8	1.06	(0.99, 1.01)	13.6		
13 item13	0.004	0.005	0.99	(0.99, 1.01)	-2.7	0.99	(1.00, 1.00)	-7.0		
14 item14	-0.534	0.005	0.97	(0.99, 1.01)	-7.1	0.97	(1.00, 1.00)	-15.1		
15 item15	-0.367	0.005	0.95	(0.99, 1.01)	-10.3	0.96	(1.00, 1.00)	-25.1		
16 item16	-0.083	0.005	0.90	(0.99, 1.01)	-22.7	0.90	(1.00, 1.00)	-53.1		
17 item17	0.280	0.005	1.01	(0.99, 1.01)	2.4	1.00	(1.00, 1.00)	-1.1		
18 item18	-0.498	0.005	0.96	(0.99, 1.01)	-8.0	0.97	(1.00, 1.00)	-18.0		
19 item19	-0.583	0.005	0.98	(0.99, 1.01)	-4.9	0.98	(1.00, 1.00)	-9.7		
20 item20	1.151	0.005	1.11	(0.99, 1.01)	22.3	1.02	(0.99, 1.01)	4.9		
21 item21	-1.040	0.005	0.94	(0.99, 1.01)	-12.1	0.96	(0.99, 1.01)	-14.8		
22 item22	-0.370	0.005	0.95	(0.99, 1.01)	-11.4	0.95	(1.00, 1.00)	-27.0		

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
23	item23	0.468	0.005	1.05	(0.99, 1.01)	9.4	1.03	(0.99, 1.01)	9.9
24	item24	0.103	0.005	1.00	(0.99, 1.01)	-0.9	0.99	(1.00, 1.00)	-3.9
25	item25	0.525	0.005	1.07	(0.99, 1.01)	14.7	1.05	(0.99, 1.01)	15.6
26	item26	0.836	0.005	1.10	(0.99, 1.01)	20.6	1.05	(0.99, 1.01)	13.2
27	item27	-0.420	0.005	0.91	(0.99, 1.01)	-20.2	0.92	(1.00, 1.00)	-46.9
28	item28	-0.644	0.005	0.92	(0.99, 1.01)	-17.1	0.94	(1.00, 1.00)	-32.5
29	item29	0.485	0.005	1.07	(0.99, 1.01)	14.4	1.05	(0.99, 1.01)	15.8
30	item30	-0.014	0.005	0.98	(0.99, 1.01)	-4.9	0.98	(1.00, 1.00)	-12.8
31	item31	0.341	0.005	1.08	(0.99, 1.01)	15.8	1.06	(0.99, 1.01)	22.9
32	item32	0.059	0.005	1.02	(0.99, 1.01)	4.9	1.02	(1.00, 1.00)	8.8
33	item33	-0.219	0.004	1.00	(0.99, 1.01)	-0.0	1.00	(0.99, 1.01)	0.0
34	item34	0.762	0.004	1.13	(0.99, 1.01)	25.5	1.11	(0.99, 1.01)	26.8
35	item35	0.521	0.004	1.08	(0.99, 1.01)	16.6	1.08	(0.99, 1.01)	18.6
36	item36	0.619	0.004	1.16	(0.99, 1.01)	32.4	1.14	(0.99, 1.01)	32.9
37	item37	1.014*	0.028	1.12	(0.99, 1.01)	24.1	1.11	(0.99, 1.01)	26.1

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 560262.40, df = 36, sig Level = 0.000

[^] Quick standard errors have been used

TERM 2: item*step

VARIABLES				UNWEIGHTED FIT				WEIGHTED FIT				
item	step	ESTIMATE	ERROR^	MNSQ	CI		T	MNSQ	CI		T	
33	item33	0		1.11	(0.99,	1.01)	21.5	1.07	(0.99,	1.01)	21.0	
33	item33	1	-0.498	0.007	1.00	(0.99,	1.01)	0.4	1.00	(1.00,	1.00)	1.0
33	item33	2	0.498*		0.89	(0.99,	1.01)	-24.0	0.93	(0.99,	1.01)	-19.8
34	item34	0		1.12	(0.99,	1.01)	23.5	1.10	(1.00,	1.00)	48.7	
34	item34	1	-1.082	0.007	1.06	(0.99,	1.01)	11.9	1.05	(1.00,	1.00)	37.2
34	item34	2	1.082*		1.29	(0.99,	1.01)	56.0	1.05	(0.98,	1.02)	4.7
35	item35	0		1.13	(0.99,	1.01)	25.3	1.11	(1.00,	1.00)	50.4	
35	item35	1	-0.614	0.007	1.07	(0.99,	1.01)	15.1	1.06	(1.00,	1.00)	31.2
35	item35	2	0.614*		0.97	(0.99,	1.01)	-7.1	0.98	(0.99,	1.01)	-2.2
36	item36	0		1.09	(0.99,	1.01)	18.6	1.07	(1.00,	1.00)	34.0	
36	item36	1	-1.096	0.007	1.01	(0.99,	1.01)	1.9	1.01	(1.00,	1.00)	7.0
36	item36	2	1.096*		1.69	(0.99,	1.01)	120.5	1.13	(0.98,	1.02)	12.7
37	item37	0		1.08	(0.99,	1.01)	15.6	1.07	(1.00,	1.00)	33.6	
37	item37	1	-1.378	0.007	1.03	(0.99,	1.01)	5.9	1.03	(1.00,	1.00)	17.9
37	item37	2	1.378*		1.72	(0.99,	1.01)	124.8	1.11	(0.97,	1.03)	7.1

An asterisk next to a parameter estimate indicates that it is constrained

[^] Quick standard errors have been used

B3.2D โมเดลพหุมิติระหว่างข้อสอบที่ให้คะแนนแบบสองค่า

VARIABLES				UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T		
1	item1	-0.765	0.006	1.12	(0.99, 1.01)	24.4	1.08	(0.99, 1.01)	25.7		
2	item2	-1.025	0.006	0.91	(0.99, 1.01)	-19.8	0.93	(0.99, 1.01)	-25.6		
3	item3	-0.055	0.006	1.09	(0.99, 1.01)	19.0	1.07	(1.00, 1.00)	28.6		
4	item4	-0.460	0.006	0.95	(0.99, 1.01)	-11.4	0.96	(1.00, 1.00)	-20.1		
5	item5	-0.444	0.006	1.01	(0.99, 1.01)	2.7	1.01	(1.00, 1.00)	3.9		
6	item6	-0.882	0.006	0.92	(0.99, 1.01)	-16.6	0.94	(1.00, 1.00)	-24.4		
7	item7	-1.057	0.007	0.95	(0.99, 1.01)	-11.7	0.96	(0.99, 1.01)	-13.2		
8	item8	-0.133	0.005	1.01	(0.99, 1.01)	2.8	1.01	(1.00, 1.00)	3.3		
9	item9	0.954	0.006	1.04	(0.99, 1.01)	9.1	0.99	(0.99, 1.01)	-1.7		
10	item10	-1.075	0.005	0.98	(0.99, 1.01)	-3.5	0.98	(0.99, 1.01)	-5.8		
11	item11	0.054	0.005	1.02	(0.99, 1.01)	5.1	1.01	(0.99, 1.01)	5.0		
12	item12	0.887	0.005	1.22	(0.99, 1.01)	43.9	1.13	(0.99, 1.01)	28.0		
13	item13	-0.078	0.005	1.02	(0.99, 1.01)	4.1	1.01	(1.00, 1.00)	4.5		
14	item14	-0.737	0.006	1.00	(0.99, 1.01)	0.1	1.00	(0.99, 1.01)	0.7		
15	item15	-0.456	0.005	0.97	(0.99, 1.01)	-6.5	0.97	(1.00, 1.00)	-13.8		
16	item16	-0.153*	0.014	0.90	(0.99, 1.01)	-20.8	0.92	(1.00, 1.00)	-37.0		
17	item17	0.161	0.007	1.12	(0.99, 1.01)	23.5	1.06	(0.99, 1.01)	19.4		
18	item18	-0.695	0.006	0.99	(0.99, 1.01)	-2.2	0.99	(0.99, 1.01)	-3.3		
19	item19	-0.817	0.006	1.01	(0.99, 1.01)	1.8	1.01	(1.00, 1.00)	2.9		
20	item20	1.055	0.007	1.35	(0.99, 1.01)	65.6	1.11	(0.99, 1.01)	21.1		
21	item21	-1.327	0.006	0.95	(0.99, 1.01)	-11.3	0.96	(0.99, 1.01)	-13.0		
22	item22	-0.613	0.006	0.97	(0.99, 1.01)	-6.8	0.97	(1.00, 1.00)	-10.6		
23	item23	0.317	0.005	1.07	(0.99, 1.01)	14.9	1.04	(0.99, 1.01)	13.6		
24	item24	-0.083	0.005	1.01	(0.99, 1.01)	2.2	1.00	(0.99, 1.01)	1.7		
25	item25	0.447	0.007	1.13	(0.99, 1.01)	26.7	1.08	(0.99, 1.01)	25.3		
26	item26	0.759	0.007	1.17	(0.99, 1.01)	33.9	1.08	(0.99, 1.01)	20.6		
27	item27	-0.563	0.006	0.91	(0.99, 1.01)	-19.1	0.93	(1.00, 1.00)	-33.2		
28	item28	-1.178	0.006	0.93	(0.99, 1.01)	-14.0	0.95	(1.00, 1.00)	-21.8		
29	item29	-0.011	0.006	1.04	(0.99, 1.01)	8.3	1.03	(0.99, 1.01)	8.5		
30	item30	-0.234*	0.007	1.00	(0.99, 1.01)	-1.0	0.99	(1.00, 1.00)	-2.2		
31	item31	-0.153	0.006	1.05	(0.99, 1.01)	11.2	1.04	(0.99, 1.01)	14.8		
32	item32	-0.440	0.006	1.01	(0.99, 1.01)	2.3	1.01	(1.00, 1.00)	3.5		

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
33 item33	0.765*	0.006	0.93	(0.99, 1.01)	-14.8	0.96	(0.99, 1.01)	-8.9
34 item34	2.328*	0.013	1.23	(0.99, 1.01)	45.0	1.01	(0.98, 1.02)	0.8
35 item35	1.703*	0.013	0.96	(0.99, 1.01)	-8.0	0.99	(0.99, 1.01)	-1.8
36 item36	1.782*	0.013	1.26	(0.99, 1.01)	51.2	1.06	(0.98, 1.02)	6.1
37 item37	2.223*	0.018	1.35	(0.99, 1.01)	66.5	1.09	(0.98, 1.02)	9.1

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 401753.72, df = 30, Sig Level = 0.000

[^] Quick standard errors have been used

B3.2P โมเดลพหุมิติระหว่างข้อสอบที่ให้คะแนนแบบหลายค่า

TERM 1: item

VARIABLES			UNWEIGHTED FIT				WEIGHTED FIT			
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T		
1 item1	-0.183	0.004	1.06	(0.99, 1.01)	12.3	1.05	(1.00, 1.00)	18.1		
2 item2	-0.871	0.005	0.91	(0.99, 1.01)	-19.0	0.93	(0.99, 1.01)	-25.8		
3 item3	0.083	0.005	1.07	(0.99, 1.01)	14.2	1.05	(1.00, 1.00)	23.0		
4 item4	-0.309	0.005	0.94	(0.99, 1.01)	-11.8	0.95	(1.00, 1.00)	-23.5		
5 item5	-0.291	0.005	1.00	(0.99, 1.01)	-0.2	1.00	(1.00, 1.00)	-0.9		
6 item6	-0.720	0.005	0.92	(0.99, 1.01)	-16.5	0.94	(1.00, 1.00)	-26.0		
7 item7	-0.682	0.005	0.93	(0.99, 1.01)	-14.0	0.95	(1.00, 1.00)	-19.9		
8 item8	-0.133	0.005	1.01	(0.99, 1.01)	3.0	1.01	(1.00, 1.00)	3.9		
9 item9	0.953	0.006	1.04	(0.99, 1.01)	8.7	0.99	(0.99, 1.01)	-2.2		
10 item10	-1.073	0.005	0.99	(0.99, 1.01)	-2.9	0.99	(0.99, 1.01)	-4.1		
11 item11	0.054	0.005	1.02	(0.99, 1.01)	5.1	1.01	(0.99, 1.01)	5.0		
12 item12	0.886	0.005	1.22	(0.99, 1.01)	43.1	1.12	(0.99, 1.01)	26.5		
13 item13	-0.078	0.005	1.02	(0.99, 1.01)	4.3	1.01	(1.00, 1.00)	4.8		
14 item14	-0.378	0.005	0.98	(0.99, 1.01)	-5.2	0.98	(1.00, 1.00)	-8.9		
15 item15	-0.456	0.005	0.97	(0.99, 1.01)	-5.8	0.97	(1.00, 1.00)	-11.5		
16 item16	-0.153*	0.014	0.91	(0.99, 1.01)	-19.5	0.92	(1.00, 1.00)	-34.2		
17 item17	0.476	0.005	1.04	(0.99, 1.01)	9.2	1.02	(0.99, 1.01)	7.8		
18 item18	-0.338	0.005	0.97	(0.99, 1.01)	-6.3	0.98	(1.00, 1.00)	-11.2		
19 item19	-0.517	0.004	0.98	(0.99, 1.01)	-4.5	0.98	(1.00, 1.00)	-8.4		
20 item20	1.266	0.005	1.17	(0.99, 1.01)	33.0	1.06	(0.99, 1.01)	11.7		
21 item21	-1.001	0.004	0.94	(0.99, 1.01)	-13.8	0.95	(0.99, 1.01)	-17.5		
22 item22	-0.324	0.004	0.95	(0.99, 1.01)	-11.1	0.96	(1.00, 1.00)	-22.6		
23 item23	0.320	0.005	1.08	(0.99, 1.01)	17.1	1.05	(0.99, 1.01)	16.3		
24 item24	-0.084	0.005	1.02	(0.99, 1.01)	4.2	1.01	(0.99, 1.01)	4.7		
25 item25	0.578	0.005	1.10	(0.99, 1.01)	20.5	1.06	(0.99, 1.01)	20.1		
26 item26	0.879	0.005	1.13	(0.99, 1.01)	25.4	1.06	(0.99, 1.01)	15.7		
27 item27	-0.420	0.005	0.91	(0.99, 1.01)	-18.7	0.93	(1.00, 1.00)	-35.8		
28 item28	-0.830	0.004	0.95	(0.99, 1.01)	-11.4	0.96	(1.00, 1.00)	-20.0		
29 item29	0.308	0.005	1.02	(0.99, 1.01)	5.1	1.01	(0.99, 1.01)	5.1		
30 item30	-0.237*	0.007	1.00	(0.99, 1.01)	-0.2	1.00	(0.99, 1.01)	-0.3		
31 item31	0.170	0.004	1.03	(0.99, 1.01)	6.7	1.02	(0.99, 1.01)	9.5		
32 item32	-0.110	0.004	1.00	(0.99, 1.01)	0.4	1.00	(1.00, 1.00)	0.4		
33 item33	0.183*	0.004	1.00	(0.99, 1.01)	0.7	1.00	(0.99, 1.01)	1.1		
34 item34	0.922*	0.009	1.14	(0.99, 1.01)	27.3	1.12	(0.99, 1.01)	26.9		
35 item35	0.577*	0.009	1.08	(0.99, 1.01)	16.1	1.07	(0.99, 1.01)	16.4		
36 item36	0.463*	0.009	1.09	(0.99, 1.01)	19.0	1.08	(0.99, 1.01)	19.3		
37 item37	1.071*	0.013	1.16	(0.99, 1.01)	31.0	1.14	(0.99, 1.01)	32.7		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 470216.30, df = 30, Sig Level = 0.000

[^] Quick standard errors have been used

TERM 2: item*step

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item	step	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
33 item33	0			1.10	(0.99, 1.01)	20.0	1.06	(0.99, 1.01)	16.2
33 item33	1	-0.639	0.007	1.00	(0.99, 1.01)	0.1	1.00	(1.00, 1.00)	1.1
33 item33	2	0.639*		0.91	(0.99, 1.01)	-18.5	0.94	(0.99, 1.01)	-13.9
34 item34	0			1.15	(0.99, 1.01)	29.6	1.12	(1.00, 1.00)	47.9
34 item34	1	-1.129	0.007	1.07	(0.99, 1.01)	14.1	1.06	(1.00, 1.00)	35.5
34 item34	2	1.129*		1.32	(0.99, 1.01)	61.1	1.01	(0.98, 1.02)	1.2
35 item35	0			1.13	(0.99, 1.01)	25.4	1.10	(1.00, 1.00)	42.5
35 item35	1	-0.658	0.007	1.07	(0.99, 1.01)	14.8	1.05	(1.00, 1.00)	28.2
35 item35	2	0.658*		0.96	(0.99, 1.01)	-8.4	0.98	(0.99, 1.01)	-3.0
36 item36	0			1.06	(0.99, 1.01)	12.4	1.04	(1.00, 1.00)	19.5
36 item36	1	-1.101	0.007	1.01	(0.99, 1.01)	1.1	1.01	(1.00, 1.00)	3.9
36 item36	2	1.101*		1.39	(0.99, 1.01)	73.0	1.07	(0.98, 1.02)	7.2
37 item37	0			1.13	(0.99, 1.01)	26.2	1.10	(1.00, 1.00)	45.0
37 item37	1	-1.420	0.007	1.06	(0.99, 1.01)	11.7	1.05	(1.00, 1.00)	29.0
37 item37	2	1.420*		1.85	(0.99, 1.01)	143.5	1.08	(0.97, 1.03)	5.7

B3.3D โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนแบบสองค่าโดยข้อสอบทุกข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	-0.155	0.004	1.12	(0.99, 1.01)	24.8	1.08	(0.99, 1.01)	25.7
2 item2	-0.308	0.004	0.91	(0.99, 1.01)	-18.9	0.94	(0.99, 1.01)	-23.5
3 item3	0.176	0.004	1.09	(0.99, 1.01)	18.7	1.07	(1.00, 1.00)	28.8
4 item4	-0.026	0.004	0.95	(0.99, 1.01)	-10.9	0.96	(1.00, 1.00)	-18.8
5 item5	-0.018	0.004	1.02	(0.99, 1.01)	3.3	1.01	(1.00, 1.00)	5.4
6 item6	-0.236	0.004	0.93	(0.99, 1.01)	-15.9	0.94	(0.99, 1.01)	-22.8
7 item7	-0.264	0.004	0.94	(0.99, 1.01)	-12.3	0.96	(0.99, 1.01)	-14.5
8 item8	0.151	0.004	1.02	(0.99, 1.01)	3.6	1.01	(1.00, 1.00)	5.0
9 item9	0.699	0.004	1.06	(0.99, 1.01)	11.7	1.00	(0.99, 1.01)	0.0
10 item10	-0.322	0.004	0.99	(0.99, 1.01)	-3.0	0.99	(0.99, 1.01)	-4.6
11 item11	0.246	0.004	1.03	(0.99, 1.01)	5.9	1.02	(0.99, 1.01)	6.2
12 item12	0.665	0.004	1.24	(0.99, 1.01)	46.1	1.14	(0.99, 1.01)	29.4
13 item13	0.179	0.004	1.02	(0.99, 1.01)	4.8	1.02	(1.00, 1.00)	6.1
14 item14	-0.104	0.004	1.00	(0.99, 1.01)	0.3	1.00	(1.00, 1.00)	1.0
15 item15	-0.012	0.004	0.97	(0.99, 1.01)	-6.0	0.97	(1.00, 1.00)	-11.8
16 item16	0.141	0.004	0.91	(0.99, 1.01)	-20.6	0.92	(1.00, 1.00)	-35.5
17 item17	0.347	0.004	1.11	(0.99, 1.01)	22.8	1.06	(0.99, 1.01)	19.9
18 item18	-0.082	0.004	0.99	(0.99, 1.01)	-1.4	0.99	(1.00, 1.00)	-2.4
19 item19	-0.134	0.004	1.02	(0.99, 1.01)	3.6	1.01	(0.99, 1.01)	4.9
20 item20	0.813	0.005	1.35	(0.99, 1.01)	66.6	1.13	(0.99, 1.01)	23.5
21 item21	-0.391	0.004	0.95	(0.99, 1.01)	-9.7	0.97	(0.99, 1.01)	-11.4
22 item22	-0.030	0.004	0.97	(0.99, 1.01)	-6.5	0.98	(1.00, 1.00)	-9.7
23 item23	0.423	0.004	1.08	(0.99, 1.01)	16.5	1.05	(0.99, 1.01)	15.5
24 item24	0.220	0.004	1.02	(0.99, 1.01)	3.3	1.01	(0.99, 1.01)	3.2
25 item25	0.426	0.004	1.13	(0.99, 1.01)	25.4	1.08	(0.99, 1.01)	24.7
26 item26	0.581	0.004	1.16	(0.99, 1.01)	31.4	1.08	(0.99, 1.01)	19.3
27 item27	-0.077	0.004	0.91	(0.99, 1.01)	-19.0	0.93	(1.00, 1.00)	-31.8
28 item28	-0.192	0.004	0.94	(0.99, 1.01)	-13.8	0.95	(1.00, 1.00)	-21.0
29 item29	0.391	0.004	1.04	(0.99, 1.01)	7.4	1.02	(0.99, 1.01)	7.2
30 item30	0.144	0.004	1.00	(0.99, 1.01)	-0.4	0.99	(0.99, 1.01)	-2.0
31 item31	0.320	0.004	1.05	(0.99, 1.01)	11.1	1.04	(0.99, 1.01)	14.6
32 item32	0.177	0.004	1.01	(0.99, 1.01)	1.8	1.01	(1.00, 1.00)	2.7
33 item33	0.615	0.004	0.93	(0.99, 1.01)	-15.2	0.95	(0.99, 1.01)	-12.1
34 item34	1.436	0.007	1.25	(0.99, 1.01)	48.5	1.03	(0.98, 1.02)	2.4
35 item35	1.141	0.005	0.96	(0.99, 1.01)	-9.5	0.98	(0.99, 1.01)	-3.0
36 item36	1.285	0.006	1.25	(0.99, 1.01)	48.9	1.05	(0.98, 1.02)	5.5
37 item37	1.306	0.006	1.32	(0.99, 1.01)	60.7	1.06	(0.98, 1.02)	6.2

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 416369.21, df = 37, sig Level = 0.000

[^] Quick standard errors have been used

B3.3P โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนแบบหลายค่าโดยข้อสอบทุกข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	-0.144	0.004	1.07	(0.99, 1.01)	14.2	1.05	(0.99, 1.01)	20.1
2 item2	-0.300	0.004	0.91	(0.99, 1.01)	-18.7	0.94	(0.99, 1.01)	-25.3
3 item3	0.175	0.004	1.06	(0.99, 1.01)	12.5	1.05	(1.00, 1.00)	21.0
4 item4	-0.023	0.004	0.94	(0.99, 1.01)	-12.4	0.95	(1.00, 1.00)	-24.4
5 item5	-0.015	0.004	1.00	(0.99, 1.01)	-0.5	1.00	(1.00, 1.00)	-1.5
6 item6	-0.230	0.004	0.92	(0.99, 1.01)	-16.4	0.94	(1.00, 1.00)	-25.8
7 item7	-0.249	0.004	0.93	(0.99, 1.01)	-14.4	0.95	(1.00, 1.00)	-21.0
8 item8	0.154	0.004	1.02	(0.99, 1.01)	5.1	1.02	(1.00, 1.00)	7.2
9 item9	0.702	0.004	1.06	(0.99, 1.01)	11.9	1.00	(0.99, 1.01)	0.7
10 item10	-0.320	0.004	0.99	(0.99, 1.01)	-2.2	0.99	(0.99, 1.01)	-3.3
11 item11	0.249	0.004	1.03	(0.99, 1.01)	6.7	1.02	(0.99, 1.01)	7.7
12 item12	0.668	0.004	1.24	(0.99, 1.01)	47.0	1.13	(0.99, 1.01)	29.2
13 item13	0.181	0.004	1.03	(0.99, 1.01)	5.8	1.02	(0.99, 1.01)	7.2
14 item14	-0.097	0.004	0.97	(0.99, 1.01)	-5.3	0.98	(1.00, 1.00)	-9.2
15 item15	-0.009	0.004	0.98	(0.99, 1.01)	-4.6	0.98	(1.00, 1.00)	-9.0
16 item16	0.144	0.004	0.91	(0.99, 1.01)	-19.5	0.92	(1.00, 1.00)	-32.5
17 item17	0.330	0.004	1.05	(0.99, 1.01)	9.4	1.03	(0.99, 1.01)	8.9
18 item18	-0.077	0.004	0.97	(0.99, 1.01)	-6.5	0.97	(1.00, 1.00)	-11.8
19 item19	-0.123	0.004	0.99	(0.99, 1.01)	-2.6	0.99	(1.00, 1.00)	-4.9
20 item20	0.777	0.004	1.18	(0.99, 1.01)	36.5	1.07	(0.99, 1.01)	14.0
21 item21	-0.368	0.004	0.94	(0.99, 1.01)	-12.0	0.96	(0.99, 1.01)	-15.1
22 item22	-0.025	0.004	0.95	(0.99, 1.01)	-9.9	0.96	(1.00, 1.00)	-18.3
23 item23	0.424	0.004	1.09	(0.99, 1.01)	17.5	1.05	(0.99, 1.01)	16.3
24 item24	0.221	0.004	1.02	(0.99, 1.01)	3.8	1.01	(0.99, 1.01)	4.0
25 item25	0.420	0.004	1.09	(0.99, 1.01)	17.7	1.06	(0.99, 1.01)	18.4
26 item26	0.573	0.004	1.11	(0.99, 1.01)	21.9	1.05	(0.99, 1.01)	14.0
27 item27	-0.074	0.004	0.91	(0.99, 1.01)	-19.1	0.93	(1.00, 1.00)	-36.9
28 item28	-0.185	0.004	0.94	(0.99, 1.01)	-12.4	0.96	(1.00, 1.00)	-21.9
29 item29	0.381	0.004	1.02	(0.99, 1.01)	3.5	1.01	(0.99, 1.01)	3.6
30 item30	0.145	0.004	1.00	(0.99, 1.01)	-0.8	1.00	(0.99, 1.01)	-1.7

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
31	item31	0.312	0.004	1.03	(0.99, 1.01)	6.1	1.02	(1.00, 1.00)	9.0
32	item32	0.173	0.004	1.00	(0.99, 1.01)	-0.3	1.00	(1.00, 1.00)	-1.1
33	item33	0.045	0.003	1.01	(0.99, 1.01)	2.1	1.01	(0.99, 1.01)	1.6
34	item34	0.554	0.003	1.14	(0.99, 1.01)	28.1	1.12	(0.99, 1.01)	28.3
35	item35	0.436	0.003	1.09	(0.99, 1.01)	18.8	1.08	(0.99, 1.01)	18.7
36	item36	0.458	0.003	1.09	(0.99, 1.01)	17.5	1.08	(0.99, 1.01)	18.3
37	item37	0.672	0.003	1.15	(0.99, 1.01)	29.6	1.13	(0.99, 1.01)	31.7

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 347494.93, df = 37, sig Level = 0.000

[^] Quick standard errors have been used

TERM 2: item*step

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item	step	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
33 item33	0			1.13	(0.99, 1.01)	25.7	1.07	(0.99, 1.01)	19.2
33 item33	1	-0.647	0.007	1.00	(0.99, 1.01)	0.7	1.00	(1.00, 1.00)	0.9
33 item33	2	0.647*		0.90	(0.99, 1.01)	-21.2	0.93	(0.99, 1.01)	-17.1
34 item34	0			1.15	(0.99, 1.01)	30.3	1.11	(1.00, 1.00)	47.6
34 item34	1	-1.142	0.007	1.07	(0.99, 1.01)	13.8	1.06	(1.00, 1.00)	35.1
34 item34	2	1.142*		1.34	(0.99, 1.01)	65.3	1.03	(0.98, 1.02)	2.8
35 item35	0			1.15	(0.99, 1.01)	30.9	1.11	(1.00, 1.00)	45.4
35 item35	1	-0.680	0.007	1.07	(0.99, 1.01)	15.3	1.06	(1.00, 1.00)	28.5
35 item35	2	0.680*		0.99	(0.99, 1.01)	-3.0	0.98	(0.99, 1.01)	-3.2
36 item36	0			1.06	(0.99, 1.01)	11.4	1.04	(1.00, 1.00)	18.2
36 item36	1	-1.092	0.007	1.01	(0.99, 1.01)	1.5	1.01	(1.00, 1.00)	5.5
36 item36	2	1.092*		1.32	(0.99, 1.01)	61.7	1.07	(0.98, 1.02)	7.7
37 item37	0			1.12	(0.99, 1.01)	25.0	1.10	(1.00, 1.00)	43.8
37 item37	1	-1.417	0.007	1.05	(0.99, 1.01)	11.0	1.05	(1.00, 1.00)	27.7
37 item37	2	1.417*		1.77	(0.99, 1.01)	131.7	1.08	(0.97, 1.03)	5.4

B3.4D โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนแบบสองค่าโดยข้อสอบบางข้อวัดสาระที่ 8

VARIABLES			UNWEIGHTED FIT				WEIGHTED FIT			
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T		
1 item1	-0.288	0.007	1.07 (0.99, 1.01)	15.2	1.06 (0.99, 1.01)	20.2				
2 item2	-0.606	0.007	0.91 (0.99, 1.01)	-18.6	0.94 (0.99, 1.01)	-23.1				
3 item3	0.185	0.004	1.15 (0.99, 1.01)	29.7	1.11 (0.99, 1.01)	37.4				
4 item4	-0.026	0.004	0.97 (0.99, 1.01)	-6.5	0.98 (0.99, 1.01)	-9.5				
5 item5	-0.030	0.007	1.01 (0.99, 1.01)	2.0	1.01 (1.00, 1.00)	2.7				
6 item6	-0.464	0.007	0.93 (0.99, 1.01)	-15.5	0.94 (0.99, 1.01)	-22.1				
7 item7	-0.280	0.004	0.97 (0.99, 1.01)	-6.7	0.98 (0.99, 1.01)	-6.1				
8 item8	0.304	0.007	1.00 (0.99, 1.01)	0.8	1.00 (1.00, 1.00)	-0.6				
9 item9	1.387	0.008	1.05 (0.99, 1.01)	9.7	0.99 (0.99, 1.01)	-1.7				
10 item10	-0.331	0.004	1.00 (0.99, 1.01)	1.0	1.01 (0.99, 1.01)	1.7				
11 item11	0.254	0.004	1.06 (0.99, 1.01)	13.2	1.03 (0.99, 1.01)	11.2				
12 item12	1.319	0.008	1.19 (0.99, 1.01)	38.0	1.09 (0.99, 1.01)	20.5				
13 item13	0.185	0.004	1.05 (0.99, 1.01)	9.7	1.03 (0.99, 1.01)	10.3				
14 item14	-0.208	0.007	1.00 (0.99, 1.01)	-1.0	1.00 (1.00, 1.00)	-1.4				
15 item15	-0.011	0.004	0.99 (0.99, 1.01)	-3.1	0.98 (1.00, 1.00)	-6.1				
16 item16	0.284	0.007	0.91 (0.99, 1.01)	-18.8	0.92 (1.00, 1.00)	-36.1				
17 item17	0.680	0.008	1.10 (0.99, 1.01)	21.2	1.04 (0.99, 1.01)	15.3				
18 item18	-0.166	0.007	0.98 (0.99, 1.01)	-3.5	0.99 (1.00, 1.00)	-5.7				
19 item19	-0.259	0.007	1.01 (0.99, 1.01)	2.0	1.01 (1.00, 1.00)	3.4				
20 item20	1.603	0.009	1.33 (0.99, 1.01)	63.3	1.08 (0.99, 1.01)	15.9				
21 item21	-0.767	0.008	0.95 (0.99, 1.01)	-9.9	0.97 (0.99, 1.01)	-9.4				
22 item22	-0.056	0.007	0.97 (0.99, 1.01)	-7.1	0.97 (1.00, 1.00)	-12.1				
23 item23	0.833	0.008	1.08 (0.99, 1.01)	15.8	1.04 (0.99, 1.01)	12.7				
24 item24	0.433	0.007	1.02 (0.99, 1.01)	3.6	1.01 (1.00, 1.00)	2.2				
25 item25	0.852	0.008	1.12 (0.99, 1.01)	24.2	1.07 (0.99, 1.01)	21.9				
26 item26	1.160	0.008	1.15 (0.99, 1.01)	29.5	1.06 (0.99, 1.01)	15.9				
27 item27	-0.079	0.004	0.93 (0.99, 1.01)	-14.8	0.95 (0.99, 1.01)	-20.6				
28 item28	-0.375	0.007	0.93 (0.99, 1.01)	-14.0	0.95 (1.00, 1.00)	-22.3				
29 item29	0.773	0.007	1.03 (0.99, 1.01)	6.8	1.02 (0.99, 1.01)	5.9				
30 item30	0.282	0.007	1.00 (0.99, 1.01)	0.0	1.00 (1.00, 1.00)	-1.7				
31 item31	0.633	0.007	1.05 (0.99, 1.01)	10.1	1.04 (0.99, 1.01)	13.3				
32 item32	0.351	0.007	1.01 (0.99, 1.01)	1.3	1.00 (1.00, 1.00)	1.6				
33 item33	1.158	0.008	0.90 (0.99, 1.01)	-22.3	0.94 (0.99, 1.01)	-16.6				
34 item34	2.852	0.013	1.22 (0.99, 1.01)	43.7	1.03 (0.98, 1.02)	2.4				
35 item35	2.250	0.011	0.94 (0.99, 1.01)	-11.8	0.98 (0.99, 1.01)	-3.1				
36 item36	2.540	0.012	1.22 (0.99, 1.01)	43.4	1.04 (0.98, 1.02)	4.3				
37 item37	2.605	0.012	1.29 (0.99, 1.01)	56.5	1.06 (0.98, 1.02)	5.7				

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 412227.58, df = 37, sig Level = 0.000

[^] Quick standard errors have been used

B3.4P โมเดลพหุมิติภายในข้อสอบที่ให้คะแนนแบบหลายค่าโดยข้อสอบบางข้อวัดสาระที่ 8

TERM 1: item									
VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T	
1 item1	-0.275	0.007	1.03	(0.99, 1.01)	7.2	1.03	(1.00, 1.00)	12.5	
2 item2	-0.590	0.007	0.92	(0.99, 1.01)	-18.2	0.94	(0.99, 1.01)	-24.7	
3 item3	0.182	0.004	1.13	(0.99, 1.01)	26.4	1.08	(1.00, 1.00)	32.7	
4 item4	-0.025	0.004	0.96	(0.99, 1.01)	-7.6	0.97	(1.00, 1.00)	-12.1	
5 item5	-0.028	0.007	0.99	(0.99, 1.01)	-1.3	0.99	(1.00, 1.00)	-3.8	
6 item6	-0.451	0.007	0.93	(0.99, 1.01)	-15.5	0.94	(1.00, 1.00)	-24.5	
7 item7	-0.264	0.004	0.95	(0.99, 1.01)	-10.9	0.96	(0.99, 1.01)	-12.9	
8 item8	0.299	0.007	1.00	(0.99, 1.01)	-0.7	0.99	(1.00, 1.00)	-3.2	
9 item9	1.371	0.008	1.02	(0.99, 1.01)	4.8	0.98	(0.99, 1.01)	-3.5	
10 item10	-0.332	0.004	1.01	(0.99, 1.01)	2.6	1.01	(0.99, 1.01)	3.1	
11 item11	0.254	0.004	1.07	(0.99, 1.01)	13.9	1.03	(0.99, 1.01)	12.3	
12 item12	1.304	0.008	1.17	(0.99, 1.01)	33.3	1.08	(0.99, 1.01)	18.7	
13 item13	0.185	0.004	1.06	(0.99, 1.01)	11.4	1.03	(0.99, 1.01)	11.9	
14 item14	-0.194	0.007	0.97	(0.99, 1.01)	-6.5	0.97	(1.00, 1.00)	-12.4	
15 item15	-0.012	0.004	0.99	(0.99, 1.01)	-1.6	0.99	(1.00, 1.00)	-4.6	
16 item16	0.279	0.007	0.91	(0.99, 1.01)	-19.6	0.92	(1.00, 1.00)	-39.5	
17 item17	0.648	0.007	1.04	(0.99, 1.01)	7.4	1.01	(0.99, 1.01)	4.5	
18 item18	-0.154	0.007	0.96	(0.99, 1.01)	-8.3	0.97	(1.00, 1.00)	-16.5	
19 item19	-0.244	0.007	0.98	(0.99, 1.01)	-4.0	0.99	(1.00, 1.00)	-7.1	
20 item20	1.529	0.009	1.17	(0.99, 1.01)	33.7	1.05	(0.99, 1.01)	9.2	
21 item21	-0.726	0.007	0.94	(0.99, 1.01)	-13.0	0.96	(0.99, 1.01)	-15.5	
22 item22	-0.051	0.007	0.95	(0.99, 1.01)	-10.5	0.96	(1.00, 1.00)	-21.9	
23 item23	0.826	0.008	1.07	(0.99, 1.01)	14.4	1.03	(0.99, 1.01)	11.3	
24 item24	0.432	0.007	1.01	(0.99, 1.01)	2.2	1.00	(1.00, 1.00)	0.5	
25 item25	0.832	0.008	1.08	(0.99, 1.01)	16.2	1.05	(0.99, 1.01)	14.8	
26 item26	1.134	0.008	1.09	(0.99, 1.01)	19.2	1.04	(0.99, 1.01)	10.1	
27 item27	-0.078	0.004	0.93	(0.99, 1.01)	-15.1	0.95	(1.00, 1.00)	-22.5	
28 item28	-0.364	0.007	0.94	(0.99, 1.01)	-12.9	0.95	(1.00, 1.00)	-24.7	
29 item29	0.755	0.007	1.02	(0.99, 1.01)	3.9	1.01	(0.99, 1.01)	3.3	
30 item30	0.284	0.007	1.00	(0.99, 1.01)	-0.8	0.99	(1.00, 1.00)	-3.2	
31 item31	0.618	0.007	1.03	(0.99, 1.01)	5.8	1.02	(1.00, 1.00)	8.2	
32 item32	0.343	0.007	1.00	(0.99, 1.01)	-1.0	0.99	(1.00, 1.00)	-2.8	
33 item33	0.087	0.005	1.00	(0.99, 1.01)	-0.5	1.00	(0.99, 1.01)	-1.1	
34 item34	1.079	0.006	1.12	(0.99, 1.01)	23.3	1.10	(0.99, 1.01)	22.4	
35 item35	0.848	0.005	1.09	(0.99, 1.01)	17.8	1.07	(0.99, 1.01)	16.6	
36 item36	0.899	0.006	1.08	(0.99, 1.01)	16.4	1.07	(0.99, 1.01)	16.1	
37 item37	1.327	0.006	1.13	(0.99, 1.01)	25.9	1.11	(0.99, 1.01)	27.1	

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 342035.24, df = 37, Sig Level = 0.000

[^] Quick standard errors have been used

TERM 2: item*step									
VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item	step	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
33 item33	0			1.09	(0.99, 1.01)	18.3	1.07	(0.99, 1.01)	18.2
33 item33	1	-0.575	0.007	1.01	(0.99, 1.01)	1.2	1.00	(1.00, 1.00)	1.8
33 item33	2	0.575*		0.89	(0.99, 1.01)	-24.2	0.93	(0.99, 1.01)	-20.2
34 item34	0			1.12	(0.99, 1.01)	23.3	1.09	(1.00, 1.00)	41.7
34 item34	1	-1.108	0.007	1.06	(0.99, 1.01)	12.6	1.05	(1.00, 1.00)	33.5
34 item34	2	1.108*		1.28	(0.99, 1.01)	53.4	1.02	(0.98, 1.02)	1.9
35 item35	0			1.13	(0.99, 1.01)	25.7	1.10	(1.00, 1.00)	43.6
35 item35	1	-0.651	0.007	1.08	(0.99, 1.01)	16.4	1.05	(1.00, 1.00)	29.5
35 item35	2	0.651*		0.97	(0.99, 1.01)	-7.2	0.98	(0.99, 1.01)	-3.2
36 item36	0			1.04	(0.99, 1.01)	8.2	1.03	(1.00, 1.00)	15.6
36 item36	1	-1.068	0.007	1.00	(0.99, 1.01)	0.1	1.00	(1.00, 1.00)	0.2
36 item36	2	1.068*		1.35	(0.99, 1.01)	67.0	1.06	(0.98, 1.02)	6.4
37 item37	0			1.11	(0.99, 1.01)	22.3	1.08	(1.00, 1.00)	38.1
37 item37	1	-1.395	0.007	1.04	(0.99, 1.01)	9.3	1.04	(1.00, 1.00)	24.3
37 item37	2	1.395*		1.70	(0.99, 1.01)	121.8	1.07	(0.97, 1.03)	5.1

An asterisk next to a parameter estimate indicates that it is constrained

[^] Quick standard errors have been used

ตารางที่ C1 ค่าประมาณพารามิเตอร์ความยากข้อสอบสำหรับกำหนดคะแนนจุดตัดขั้นต้น
ในภาพรวม

C1.1 ปีการศึกษา 2550

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	1.133	0.004	1.15 (0.99, 1.01)	42.9	1.08 (0.99, 1.01)	29.3		
2 item2	-0.557	0.004	1.01 (0.99, 1.01)	3.6	1.02 (1.00, 1.00)	10.7		
3 item3	-0.182	0.004	1.01 (0.99, 1.01)	4.3	1.01 (1.00, 1.00)	7.4		
4 item4	0.627	0.004	1.11 (0.99, 1.01)	30.8	1.07 (1.00, 1.00)	36.7		
5 item5	0.249	0.004	0.99 (0.99, 1.01)	-3.5	0.99 (1.00, 1.00)	-7.8		
6 item6	0.718	0.004	1.00 (0.99, 1.01)	-0.7	0.99 (1.00, 1.00)	-4.3		
7 item7	0.697	0.004	1.05 (0.99, 1.01)	14.0	1.03 (1.00, 1.00)	13.1		
8 item8	-0.743	0.004	0.93 (0.99, 1.01)	-20.7	0.95 (1.00, 1.00)	-25.6		
9 item9	-0.365	0.004	0.95 (0.99, 1.01)	-14.3	0.96 (1.00, 1.00)	-23.5		
10 item10	-0.480	0.004	0.93 (0.99, 1.01)	-20.7	0.95 (1.00, 1.00)	-27.7		
11 item11	-0.388	0.004	0.93 (0.99, 1.01)	-21.8	0.94 (1.00, 1.00)	-34.0		
12 item12	-0.361	0.004	0.93 (0.99, 1.01)	-21.8	0.95 (1.00, 1.00)	-32.4		
13 item13	-0.199	0.004	0.96 (0.99, 1.01)	-12.1	0.97 (1.00, 1.00)	-19.0		
14 item14	0.771	0.004	1.13 (0.99, 1.01)	36.4	1.08 (1.00, 1.00)	37.3		
15 item15	-0.694	0.004	0.91 (0.99, 1.01)	-27.9	0.94 (1.00, 1.00)	-33.2		
16 item16	-0.117	0.004	1.00 (0.99, 1.01)	-0.9	1.00 (1.00, 1.00)	-0.9		
17 item17	0.938	0.004	1.14 (0.99, 1.01)	40.1	1.08 (1.00, 1.00)	31.6		
18 item18	0.334	0.004	1.02 (0.99, 1.01)	7.4	1.02 (1.00, 1.00)	9.0		
19 item19	-0.325	0.004	1.02 (0.99, 1.01)	6.1	1.02 (1.00, 1.00)	10.6		
20 item20	-0.408	0.004	0.94 (0.99, 1.01)	-18.7	0.95 (1.00, 1.00)	-28.2		
21 item21	-0.305	0.004	0.93 (0.99, 1.01)	-21.3	0.95 (1.00, 1.00)	-33.3		
22 item22	-0.704	0.004	1.00 (0.99, 1.01)	-1.3	0.99 (1.00, 1.00)	-3.8		
23 item23	-0.171	0.004	1.03 (0.99, 1.01)	7.8	1.02 (1.00, 1.00)	14.5		
24 item24	0.349	0.004	0.98 (0.99, 1.01)	-6.4	0.98 (1.00, 1.00)	-15.0		
25 item25	0.669	0.004	1.07 (0.99, 1.01)	20.5	1.04 (1.00, 1.00)	19.6		
26 item26	0.159	0.004	1.02 (0.99, 1.01)	4.6	1.01 (1.00, 1.00)	7.1		
27 item27	-0.400	0.004	0.99 (0.99, 1.01)	-4.1	0.99 (1.00, 1.00)	-7.5		
28 item28	0.096	0.004	1.11 (0.99, 1.01)	31.5	1.09 (1.00, 1.00)	55.8		
29 item29	-0.673	0.004	0.92 (0.99, 1.01)	-25.9	0.95 (1.00, 1.00)	-26.5		
30 item30	0.418	0.004	1.05 (0.99, 1.01)	14.3	1.03 (1.00, 1.00)	18.2		
31 item31	-0.177	0.004	1.01 (0.99, 1.01)	2.3	1.00 (1.00, 1.00)	0.6		
32 item32	0.616	0.004	1.08 (0.99, 1.01)	22.4	1.05 (1.00, 1.00)	25.7		
33 item33	0.036	0.004	1.01 (0.99, 1.01)	2.8	1.01 (1.00, 1.00)	4.3		
34 item34	-0.014	0.004	0.96 (0.99, 1.01)	-12.6	0.96 (1.00, 1.00)	-25.0		
35 item35	-0.634	0.004	0.94 (0.99, 1.01)	-18.1	0.96 (1.00, 1.00)	-22.0		
36 item36	-0.603	0.004	0.91 (0.99, 1.01)	-27.7	0.94 (1.00, 1.00)	-34.7		
37 item37	0.052	0.004	0.98 (0.99, 1.01)	-6.3	0.98 (1.00, 1.00)	-12.1		
38 item38	0.438	0.004	0.98 (0.99, 1.01)	-6.9	0.97 (1.00, 1.00)	-17.2		
39 item39	-0.661	0.004	0.93 (0.99, 1.01)	-21.6	0.96 (1.00, 1.00)	-23.4		
40 item40	0.862*	0.023	1.14 (0.99, 1.01)	40.7	1.09 (1.00, 1.00)	37.6		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 746838.40, df = 39, Sig Level = 0.000

[^] Quick standard errors have been used

C1.2 ปีการศึกษา 2551

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	0.515	0.004	1.13 (0.99, 1.01)	35.7	1.10 (1.00, 1.00)	53.0		
2 item2	0.718	0.004	0.99 (0.99, 1.01)	-1.9	0.97 (1.00, 1.00)	-12.6		
3 item3	0.926	0.004	1.06 (0.99, 1.01)	15.2	1.02 (1.00, 1.00)	6.5		
4 item4	0.197	0.004	1.04 (0.99, 1.01)	10.7	1.03 (1.00, 1.00)	18.0		
5 item5	0.835	0.004	1.08 (0.99, 1.01)	21.0	1.05 (1.00, 1.00)	20.2		
6 item6	-0.223	0.004	0.96 (0.99, 1.01)	-10.5	0.97 (1.00, 1.00)	-15.0		
7 item7	-0.972	0.004	0.84 (0.99, 1.01)	-46.6	0.90 (0.99, 1.01)	-36.2		
8 item8	-0.471	0.004	0.95 (0.99, 1.01)	-13.6	0.97 (1.00, 1.00)	-17.7		
9 item9	0.469	0.004	0.99 (0.99, 1.01)	-2.9	0.98 (1.00, 1.00)	-11.1		
10 item10	0.198	0.004	1.06 (0.99, 1.01)	15.7	1.05 (1.00, 1.00)	28.2		
11 item11	-0.192	0.004	1.01 (0.99, 1.01)	3.9	1.01 (1.00, 1.00)	7.2		
12 item12	-0.151	0.004	0.96 (0.99, 1.01)	-10.4	0.97 (1.00, 1.00)	-14.9		
13 item13	-0.381	0.004	0.90 (0.99, 1.01)	-28.8	0.92 (1.00, 1.00)	-42.9		
14 item14	-0.010	0.004	1.00 (0.99, 1.01)	0.7	1.00 (1.00, 1.00)	1.7		
15 item15	-1.447	0.004	0.74 (0.99, 1.01)	-79.2	0.86 (0.99, 1.01)	-39.7		

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
16 item16	-0.261	0.004	1.02	(0.99, 1.01)	4.5	1.01	(1.00, 1.00)	6.4
17 item17	-0.680	0.004	0.91	(0.99, 1.01)	-25.0	0.94	(1.00, 1.00)	-27.8
18 item18	0.125	0.004	0.99	(0.99, 1.01)	-3.9	0.99	(1.00, 1.00)	-8.8
19 item19	-0.334	0.004	0.99	(0.99, 1.01)	-3.6	0.99	(1.00, 1.00)	-3.5
20 item20	-0.363	0.004	0.97	(0.99, 1.01)	-7.8	0.98	(1.00, 1.00)	-12.1
21 item21	0.738	0.004	1.02	(0.99, 1.01)	6.1	0.99	(1.00, 1.00)	-3.5
22 item22	-0.878	0.004	0.89	(0.99, 1.01)	-30.8	0.93	(0.99, 1.01)	-26.4
23 item23	1.102	0.004	1.11	(0.99, 1.01)	29.6	1.06	(0.99, 1.01)	21.3
24 item24	0.496	0.004	1.01	(0.99, 1.01)	3.9	1.00	(1.00, 1.00)	0.8
25 item25	0.466	0.004	1.06	(0.99, 1.01)	15.2	1.04	(1.00, 1.00)	21.9
26 item26	-0.606	0.004	0.93	(0.99, 1.01)	-18.8	0.96	(1.00, 1.00)	-19.4
27 item27	-0.702	0.004	1.04	(0.99, 1.01)	10.3	1.02	(1.00, 1.00)	8.1
28 item28	0.712	0.004	1.01	(0.99, 1.01)	3.0	0.99	(1.00, 1.00)	-6.4
29 item29	-0.323	0.004	1.02	(0.99, 1.01)	4.4	1.00	(1.00, 1.00)	1.6
30 item30	-0.538	0.004	0.84	(0.99, 1.01)	-46.4	0.89	(1.00, 1.00)	-55.8
31 item31	0.458	0.004	1.07	(0.99, 1.01)	18.9	1.05	(1.00, 1.00)	28.2
32 item32	0.458	0.004	1.07	(0.99, 1.01)	18.6	1.05	(1.00, 1.00)	26.7
33 item33	0.596	0.004	1.05	(0.99, 1.01)	12.9	1.03	(1.00, 1.00)	16.6
34 item34	-0.779	0.004	0.92	(0.99, 1.01)	-22.0	0.96	(1.00, 1.00)	-18.0
35 item35	0.411	0.004	1.10	(0.99, 1.01)	26.6	1.08	(1.00, 1.00)	45.1
36 item36	0.220	0.004	1.02	(0.99, 1.01)	4.3	1.01	(1.00, 1.00)	6.6
37 item37	-0.465	0.004	0.98	(0.99, 1.01)	-6.4	0.99	(1.00, 1.00)	-6.3
38 item38	-0.067	0.004	0.93	(0.99, 1.01)	-21.4	0.94	(1.00, 1.00)	-36.0
39 item39	-0.302	0.004	0.98	(0.99, 1.01)	-4.9	0.98	(1.00, 1.00)	-8.4
40 item40	0.503*	0.024	1.14	(0.99, 1.01)	37.8	1.11	(1.00, 1.00)	58.4

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 860209.24, df = 39, Sig Level = 0.000

[^] Quick standard errors have been used

C1.3 ปีการศึกษา 2552

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	-0.552	0.003	1.03	(0.99, 1.01)	8.2	1.03	(1.00, 1.00)	19.4
2 item2	-0.875	0.003	0.91	(0.99, 1.01)	-29.1	0.93	(1.00, 1.00)	-44.7
3 item3	0.042	0.003	1.04	(0.99, 1.01)	12.5	1.04	(1.00, 1.00)	23.9
4 item4	-0.334	0.003	0.94	(0.99, 1.01)	-19.3	0.94	(1.00, 1.00)	-45.8
5 item5	-0.328	0.003	0.99	(0.99, 1.01)	-4.0	0.99	(1.00, 1.00)	-10.2
6 item6	-0.743	0.003	0.92	(0.99, 1.01)	-25.5	0.93	(1.00, 1.00)	-46.1
7 item7	-0.783	0.003	0.92	(0.99, 1.01)	-24.8	0.94	(1.00, 1.00)	-42.6
8 item8	-0.014	0.003	0.99	(0.99, 1.01)	-4.0	0.99	(1.00, 1.00)	-9.8
9 item9	1.019	0.003	0.99	(0.99, 1.01)	-1.7	0.97	(0.99, 1.01)	-8.9
10 item10	-0.924	0.003	0.96	(0.99, 1.01)	-12.8	0.97	(1.00, 1.00)	-18.6
11 item11	0.150	0.003	0.99	(0.99, 1.01)	-3.2	0.99	(1.00, 1.00)	-9.3
12 item12	0.927	0.003	1.12	(0.99, 1.01)	33.5	1.06	(0.99, 1.01)	20.3
13 item13	0.011	0.003	0.99	(0.99, 1.01)	-4.0	0.99	(1.00, 1.00)	-10.2
14 item14	-0.531	0.003	0.97	(0.99, 1.01)	-10.5	0.97	(1.00, 1.00)	-22.2
15 item15	-0.366	0.003	0.95	(0.99, 1.01)	-14.6	0.96	(1.00, 1.00)	-35.5
16 item16	-0.078	0.003	0.89	(0.99, 1.01)	-32.9	0.90	(1.00, 1.00)	-76.7
17 item17	0.283	0.003	1.01	(0.99, 1.01)	4.2	1.00	(1.00, 1.00)	-0.2
18 item18	-0.492	0.003	0.96	(0.99, 1.01)	-11.7	0.97	(1.00, 1.00)	-26.2
19 item19	-0.588	0.003	0.98	(0.99, 1.01)	-6.8	0.98	(1.00, 1.00)	-13.5
20 item20	1.144	0.004	1.11	(0.99, 1.01)	31.0	1.02	(0.99, 1.01)	6.9
21 item21	-1.031	0.003	0.94	(0.99, 1.01)	-17.8	0.96	(1.00, 1.00)	-22.3
22 item22	-0.366	0.003	0.95	(0.99, 1.01)	-16.2	0.95	(1.00, 1.00)	-38.4
23 item23	0.468	0.003	1.05	(0.99, 1.01)	14.0	1.03	(1.00, 1.00)	14.7
24 item24	0.103	0.003	0.99	(0.99, 1.01)	-2.2	0.99	(1.00, 1.00)	-7.2
25 item25	0.526	0.003	1.07	(0.99, 1.01)	21.2	1.05	(1.00, 1.00)	22.4
26 item26	0.831	0.003	1.11	(0.99, 1.01)	30.5	1.06	(0.99, 1.01)	19.9
27 item27	-0.416	0.003	0.91	(0.99, 1.01)	-28.4	0.92	(1.00, 1.00)	-66.1
28 item28	-0.641	0.003	0.92	(0.99, 1.01)	-24.1	0.94	(1.00, 1.00)	-45.8
29 item29	0.488	0.003	1.07	(0.99, 1.01)	20.2	1.05	(1.00, 1.00)	21.5
30 item30	-0.017	0.003	0.98	(0.99, 1.01)	-7.0	0.98	(1.00, 1.00)	-18.0
31 item31	0.346	0.003	1.08	(0.99, 1.01)	22.7	1.06	(1.00, 1.00)	32.5
32 item32	0.049	0.003	1.02	(0.99, 1.01)	6.6	1.02	(1.00, 1.00)	12.2
33 item33	-0.229	0.003	1.00	(0.99, 1.01)	0.3	1.00	(1.00, 1.00)	0.4
34 item34	0.761	0.003	1.13	(0.99, 1.01)	37.1	1.12	(0.99, 1.01)	39.0
35 item35	0.518	0.003	1.08	(0.99, 1.01)	23.8	1.08	(0.99, 1.01)	26.5
36 item36	0.621	0.003	1.16	(0.99, 1.01)	46.2	1.14	(0.99, 1.01)	46.7
37 item37	1.021*	0.020	1.12	(0.99, 1.01)	34.1	1.11	(0.99, 1.01)	36.9

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 1117661.52, df = 36, Sig Level = 0.000

[^] Quick standard errors have been used

ตารางที่ C2 ค่าประมาณพารามิเตอร์ความยากข้อสอบสำหรับกำหนดคะแนนจุดตัดขั้นต้น
 รายละเอียดการเรียนรู้

C2.1 ปีการศึกษา 2550

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	0.764	0.004	1.10	(0.99, 1.01)	30.4	1.06	(0.99, 1.01)	20.4
2	item2	-0.947	0.004	1.00	(0.99, 1.01)	-1.0	1.01	(1.00, 1.00)	3.4
3	item3	-0.569	0.004	1.00	(0.99, 1.01)	-1.4	1.00	(1.00, 1.00)	-2.7
4	item4	0.246	0.004	1.07	(0.99, 1.01)	19.9	1.05	(1.00, 1.00)	23.0
5	item5	-0.137	0.004	0.98	(0.99, 1.01)	-6.8	0.98	(1.00, 1.00)	-11.5
6	item6	0.333	0.004	0.98	(0.99, 1.01)	-7.1	0.98	(1.00, 1.00)	-10.7
7	item7	0.311*	0.009	1.02	(0.99, 1.01)	5.6	1.01	(1.00, 1.00)	4.2
8	item8	-0.219	0.004	0.99	(0.99, 1.01)	-2.8	0.99	(1.00, 1.00)	-4.9
9	item9	0.219*	0.004	1.01	(0.99, 1.01)	1.6	1.00	(1.00, 1.00)	1.1
10	item10	-0.453	0.004	0.95	(0.99, 1.01)	-15.8	0.97	(1.00, 1.00)	-18.1
11	item11	-0.359	0.004	0.94	(0.99, 1.01)	-18.1	0.96	(1.00, 1.00)	-25.0
12	item12	-0.331	0.004	0.94	(0.99, 1.01)	-19.6	0.96	(1.00, 1.00)	-25.1
13	item13	-0.162	0.004	0.97	(0.99, 1.01)	-9.7	0.98	(1.00, 1.00)	-13.1
14	item14	0.835	0.004	1.14	(0.99, 1.01)	39.0	1.08	(1.00, 1.00)	37.6
15	item15	-0.657	0.004	0.92	(0.99, 1.01)	-25.5	0.94	(1.00, 1.00)	-27.5
16	item16	-0.058	0.004	1.01	(0.99, 1.01)	2.3	1.01	(1.00, 1.00)	4.6
17	item17	1.026	0.004	1.17	(0.99, 1.01)	47.2	1.09	(1.00, 1.00)	35.2
18	item18	0.415	0.004	1.04	(0.99, 1.01)	11.1	1.02	(1.00, 1.00)	13.1
19	item19	-0.255*	0.011	1.03	(0.99, 1.01)	8.4	1.02	(1.00, 1.00)	13.1
20	item20	0.065	0.004	1.00	(0.99, 1.01)	-0.5	1.00	(1.00, 1.00)	0.2
21	item21	0.190	0.004	0.98	(0.99, 1.01)	-5.5	0.99	(1.00, 1.00)	-2.4
22	item22	-0.255*	0.005	1.08	(0.99, 1.01)	24.2	1.03	(1.00, 1.00)	14.1
23	item23	-0.179	0.004	1.02	(0.99, 1.01)	5.2	1.02	(1.00, 1.00)	9.7
24	item24	0.353	0.004	0.99	(0.99, 1.01)	-3.3	0.99	(1.00, 1.00)	-7.8
25	item25	0.682	0.004	1.09	(0.99, 1.01)	24.8	1.05	(1.00, 1.00)	23.9
26	item26	0.158	0.004	1.02	(0.99, 1.01)	4.9	1.01	(1.00, 1.00)	7.6
27	item27	-0.412	0.004	0.97	(0.99, 1.01)	-8.0	0.98	(1.00, 1.00)	-13.2
28	item28	0.089	0.004	1.10	(0.99, 1.01)	29.8	1.08	(1.00, 1.00)	48.0
29	item29	-0.690*	0.009	0.91	(0.99, 1.01)	-28.6	0.94	(1.00, 1.00)	-31.8
30	item30	0.395	0.004	1.06	(0.99, 1.01)	17.3	1.04	(1.00, 1.00)	20.5
31	item31	-0.221	0.004	1.01	(0.99, 1.01)	3.0	1.00	(1.00, 1.00)	1.6
32	item32	0.596	0.004	1.09	(0.99, 1.01)	25.5	1.06	(1.00, 1.00)	27.1
33	item33	-0.006	0.004	1.02	(0.99, 1.01)	6.7	1.02	(1.00, 1.00)	10.1
34	item34	-0.060	0.004	0.98	(0.99, 1.01)	-7.4	0.98	(1.00, 1.00)	-12.6
35	item35	-0.704*	0.008	0.94	(0.99, 1.01)	-16.9	0.96	(1.00, 1.00)	-18.3
36	item36	-0.654	0.004	0.93	(0.99, 1.01)	-21.1	0.96	(1.00, 1.00)	-20.2
37	item37	0.043	0.004	1.01	(0.99, 1.01)	2.1	1.00	(1.00, 1.00)	2.2
38	item38	0.450	0.004	1.02	(0.99, 1.01)	4.8	1.00	(1.00, 1.00)	-0.5
39	item39	-0.736	0.004	0.96	(0.99, 1.01)	-11.9	0.98	(1.00, 1.00)	-7.3
40	item40	0.898*	0.008	1.22	(0.99, 1.01)	61.9	1.12	(1.00, 1.00)	48.5

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 538297.07, df = 33, Sig Level = 0.000

[^] Quick standard errors have been used

C2.2 ปีการศึกษา 2551

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item		ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1	item1	0.250	0.004	1.13	(0.99, 1.01)	35.9	1.10	(1.00, 1.00)	47.8
2	item2	0.455	0.004	1.00	(0.99, 1.01)	-0.2	0.98	(1.00, 1.00)	-7.3
3	item3	0.663	0.004	1.06	(0.99, 1.01)	17.6	1.02	(1.00, 1.00)	9.8
4	item4	-0.091	0.004	1.03	(0.99, 1.01)	9.3	1.02	(1.00, 1.00)	13.7
5	item5	0.557	0.004	1.08	(0.99, 1.01)	21.1	1.05	(1.00, 1.00)	19.6
6	item6	-0.531	0.004	0.96	(0.99, 1.01)	-11.3	0.97	(1.00, 1.00)	-15.4
7	item7	-1.302*	0.010	0.85	(0.99, 1.01)	-43.5	0.91	(0.99, 1.01)	-34.0
8	item8	-0.531	0.004	0.99	(0.99, 1.01)	-2.3	0.99	(1.00, 1.00)	-4.7
9	item9	0.531*	0.004	1.05	(0.99, 1.01)	12.6	1.02	(1.00, 1.00)	9.6
10	item10	0.565	0.004	1.10	(0.99, 1.01)	27.9	1.08	(1.00, 1.00)	40.2
11	item11	0.149	0.004	1.05	(0.99, 1.01)	14.9	1.04	(1.00, 1.00)	20.7
12	item12	0.181	0.004	0.99	(0.99, 1.01)	-2.0	1.00	(1.00, 1.00)	0.9
13	item13	-0.068	0.004	0.92	(0.99, 1.01)	-22.7	0.94	(1.00, 1.00)	-28.5
14	item14	0.314	0.004	1.04	(0.99, 1.01)	10.5	1.03	(1.00, 1.00)	15.0
15	item15	-1.222	0.004	0.77	(0.99, 1.01)	-68.8	0.90	(0.99, 1.01)	-25.4
16	item16	0.045	0.004	1.06	(0.99, 1.01)	16.8	1.04	(1.00, 1.00)	19.0
17	item17	-0.395	0.004	0.95	(0.99, 1.01)	-13.8	0.97	(1.00, 1.00)	-12.7
18	item18	0.455	0.004	1.02	(0.99, 1.01)	4.8	1.01	(1.00, 1.00)	3.2
19	item19	-0.024*	0.012	1.03	(0.99, 1.01)	8.3	1.03	(1.00, 1.00)	12.5
20	item20	-0.226	0.004	1.05	(0.99, 1.01)	13.0	1.03	(1.00, 1.00)	10.6
21	item21	1.027	0.004	1.11	(0.99, 1.01)	30.1	1.04	(1.00, 1.00)	16.5

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
22 item22	-0.800*	0.006	0.95 (0.99, 1.01)	-13.7	0.98 (0.99, 1.01)	-7.2		
23 item23	0.933	0.004	1.09 (0.99, 1.01)	25.7	1.05 (0.99, 1.01)	17.7		
24 item24	0.326	0.004	1.01 (0.99, 1.01)	1.8	1.00 (1.00, 1.00)	-0.6		
25 item25	0.301	0.004	1.05 (0.99, 1.01)	13.4	1.03 (1.00, 1.00)	17.7		
26 item26	-0.782	0.004	0.93 (0.99, 1.01)	-19.2	0.96 (1.00, 1.00)	-20.0		
27 item27	-0.873	0.004	1.02 (0.99, 1.01)	6.3	1.01 (1.00, 1.00)	3.7		
28 item28	0.570	0.004	1.01 (0.99, 1.01)	4.1	1.00 (1.00, 1.00)	-2.1		
29 item29	-0.474*	0.010	1.01 (0.99, 1.01)	2.1	1.00 (1.00, 1.00)	-2.3		
30 item30	-0.670	0.004	0.85 (0.99, 1.01)	-44.5	0.89 (1.00, 1.00)	-52.8		
31 item31	0.350	0.004	1.06 (0.99, 1.01)	16.4	1.04 (1.00, 1.00)	22.6		
32 item32	0.358	0.004	1.07 (0.99, 1.01)	19.1	1.05 (1.00, 1.00)	24.8		
33 item33	0.510	0.004	1.05 (0.99, 1.01)	13.0	1.03 (1.00, 1.00)	15.3		
34 item34	-0.879	0.004	0.91 (0.99, 1.01)	-26.4	0.94 (1.00, 1.00)	-25.4		
35 item35	0.331*	0.009	1.09 (0.99, 1.01)	25.6	1.07 (1.00, 1.00)	38.2		
36 item36	0.256	0.004	1.03 (0.99, 1.01)	7.3	1.02 (1.00, 1.00)	9.8		
37 item37	-0.459	0.004	0.97 (0.99, 1.01)	-9.6	0.98 (1.00, 1.00)	-10.2		
38 item38	-0.047	0.004	0.93 (0.99, 1.01)	-20.7	0.94 (1.00, 1.00)	-30.2		
39 item39	-0.294	0.004	0.98 (0.99, 1.01)	-5.2	0.99 (1.00, 1.00)	-7.5		
40 item40	0.545*	0.008	1.18 (0.99, 1.01)	47.3	1.13 (1.00, 1.00)	59.6		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 596478.74, df = 33, sig Level = 0.000

[^] Quick standard errors have been used

C2.3 ปีการศึกษา 2552

TERM 1: item

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 item1	-0.267	0.005	1.04 (0.99, 1.01)	11.1	1.03 (1.00, 1.00)	19.5		
2 item2	-0.592	0.005	0.92 (0.99, 1.01)	-25.5	0.94 (1.00, 1.00)	-34.4		
3 item3	0.181	0.003	1.13 (0.99, 1.01)	37.1	1.08 (1.00, 1.00)	45.9		
4 item4	-0.022	0.003	0.96 (0.99, 1.01)	-10.7	0.97 (1.00, 1.00)	-16.7		
5 item5	-0.030	0.005	0.99 (0.99, 1.01)	-1.6	0.99 (1.00, 1.00)	-4.7		
6 item6	-0.452	0.005	0.93 (0.99, 1.01)	-22.7	0.94 (1.00, 1.00)	-35.7		
7 item7	-0.266	0.003	0.95 (0.99, 1.01)	-15.9	0.96 (1.00, 1.00)	-18.4		
8 item8	0.295	0.005	1.00 (0.99, 1.01)	-0.6	0.99 (1.00, 1.00)	-3.6		
9 item9	1.373	0.006	1.03 (0.99, 1.01)	7.6	0.98 (0.99, 1.01)	-5.0		
10 item10	-0.333	0.003	1.01 (0.99, 1.01)	3.5	1.01 (1.00, 1.00)	4.6		
11 item11	0.256	0.003	1.07 (0.99, 1.01)	20.6	1.04 (1.00, 1.00)	17.9		
12 item12	1.307	0.006	1.17 (0.99, 1.01)	47.0	1.08 (0.99, 1.01)	26.5		
13 item13	0.190	0.003	1.06 (0.99, 1.01)	16.6	1.03 (1.00, 1.00)	17.5		
14 item14	-0.188	0.005	0.97 (0.99, 1.01)	-10.3	0.97 (1.00, 1.00)	-20.1		
15 item15	-0.009	0.003	0.99 (0.99, 1.01)	-2.0	0.99 (1.00, 1.00)	-5.8		
16 item16	0.287	0.005	0.91 (0.99, 1.01)	-27.7	0.92 (1.00, 1.00)	-55.2		
17 item17	0.650	0.005	1.03 (0.99, 1.01)	9.7	1.01 (1.00, 1.00)	5.0		
18 item18	-0.146	0.005	0.96 (0.99, 1.01)	-11.9	0.97 (1.00, 1.00)	-24.3		
19 item19	-0.247	0.005	0.98 (0.99, 1.01)	-5.5	0.99 (1.00, 1.00)	-9.5		
20 item20	1.525	0.006	1.17 (0.99, 1.01)	47.8	1.05 (0.99, 1.01)	12.9		
21 item21	-0.716	0.005	0.94 (0.99, 1.01)	-18.2	0.96 (1.00, 1.00)	-21.3		
22 item22	-0.045	0.005	0.95 (0.99, 1.01)	-15.3	0.96 (1.00, 1.00)	-31.7		
23 item23	0.824	0.005	1.07 (0.99, 1.01)	19.4	1.03 (1.00, 1.00)	15.2		
24 item24	0.428	0.005	1.01 (0.99, 1.01)	2.1	1.00 (1.00, 1.00)	-0.8		
25 item25	0.832	0.005	1.08 (0.99, 1.01)	22.8	1.05 (1.00, 1.00)	20.9		
26 item26	1.128	0.006	1.10 (0.99, 1.01)	28.4	1.04 (0.99, 1.01)	14.7		
27 item27	-0.076	0.003	0.93 (0.99, 1.01)	-20.7	0.95 (1.00, 1.00)	-30.7		
28 item28	-0.359	0.005	0.94 (0.99, 1.01)	-18.4	0.95 (1.00, 1.00)	-35.2		
29 item29	0.760	0.005	1.02 (0.99, 1.01)	4.8	1.01 (1.00, 1.00)	3.5		
30 item30	0.277	0.005	0.99 (0.99, 1.01)	-2.8	0.99 (1.00, 1.00)	-7.4		
31 item31	0.626	0.005	1.03 (0.99, 1.01)	8.3	1.02 (1.00, 1.00)	11.6		
32 item32	0.336	0.005	1.00 (0.99, 1.01)	-1.4	0.99 (1.00, 1.00)	-3.8		
33 item33	0.081	0.004	1.00 (0.99, 1.01)	1.1	1.01 (0.99, 1.01)	2.4		
34 item34	1.082	0.004	1.11 (0.99, 1.01)	32.6	1.09 (0.99, 1.01)	31.4		
35 item35	0.849	0.004	1.09 (0.99, 1.01)	24.8	1.07 (0.99, 1.01)	22.8		
36 item36	0.904	0.004	1.08 (0.99, 1.01)	23.6	1.07 (0.99, 1.01)	23.0		
37 item37	1.337	0.004	1.13 (0.99, 1.01)	36.6	1.11 (0.99, 1.01)	38.1		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 1.000

Chi-square test of parameter equality = 685335.59, df = 37, sig Level = 0.000

[^] Quick standard errors have been used

TERM 2: item*step

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
item	step	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
33 item33	0			1.09	(0.99, 1.01)	25.1	1.07	(0.99, 1.01)	27.5
33 item33	1	-0.582	0.005	1.01	(0.99, 1.01)	3.6	1.01	(1.00, 1.00)	9.3
33 item33	2	0.582*		0.89	(0.99, 1.01)	-32.5	0.94	(1.00, 1.00)	-24.6
34 item34	0			1.11	(0.99, 1.01)	32.2	1.09	(1.00, 1.00)	57.9
34 item34	1	-1.106	0.005	1.06	(0.99, 1.01)	16.7	1.05	(1.00, 1.00)	45.0
34 item34	2	1.106*		1.27	(0.99, 1.01)	74.3	1.02	(0.99, 1.01)	3.3
35 item35	0			1.13	(0.99, 1.01)	36.0	1.10	(1.00, 1.00)	60.4
35 item35	1	-0.656	0.005	1.08	(0.99, 1.01)	22.8	1.05	(1.00, 1.00)	40.8
35 item35	2	0.656*		0.97	(0.99, 1.01)	-8.5	0.98	(0.99, 1.01)	-4.7
36 item36	0			1.04	(0.99, 1.01)	12.1	1.03	(1.00, 1.00)	23.2
36 item36	1	-1.065	0.005	1.00	(0.99, 1.01)	0.2	1.00	(1.00, 1.00)	0.7
36 item36	2	1.065*		1.36	(0.99, 1.01)	96.6	1.06	(0.99, 1.01)	8.5
37 item37	0			1.11	(0.99, 1.01)	30.8	1.08	(1.00, 1.00)	53.2
37 item37	1	-1.397	0.005	1.04	(0.99, 1.01)	13.0	1.04	(1.00, 1.00)	33.4
37 item37	2	1.397*		1.72	(0.99, 1.01)	175.6	1.07	(0.98, 1.02)	7.1

ตารางที่ D1 ตัวอย่างคำสั่ง ConQuest

D1.1 โมเดลเอกมิตี

```
Title 10%ModelFit+polytomous+uirt+onet52;
datafile D:\#52sam10fit-3c.dat;
Format id 1-8 response 11-47;
labels <<D:\#50onet.lab;
codes 0,1,2,3,4;
recode (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
score (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
model item+item*step;
set update=yes, warnings=no;
estimate !method=montecarlo, nodes=2000, converge=.01;
show !table=1:2:3:4:5, estimates=latent>> D:\10fit-3c-uni52.shw.shw;
show cases !estimates=eap >> D:\10fit-3c-uni52.eap.eap;
itanal >> D:\10fit-3c-uni52.itn.itn;
```

D1.2 โมเดลพหุมิติภายในข้อสอบ

```
Title 10%ModelFit+polytomous+within(some8.1)+onet52;
datafile D:\#52sam10fit-3c.dat;
Format id 1-8 response 11-47;
labels <<D:\#50onet.lab;
codes 0,1,2,3,4;
recode (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
score (0,1,2,3,4) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) !item(2,5,6,25,26,37);
score (0,1,2,3,4) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(3,4,27);
score (0,1,2,3,4) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) !item(1,33);
score (0,1,2,3,4) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) !item(8,9,12,16);
score (0,1,2,3,4) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(10,11,13,15);
score (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) !item(14,17,18,34);
score (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) !item(7);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) !item(19,20,21,22,35);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) ( ) !item(23,24,30);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(28,29,31,32,36);
model item+item*step;
set constraints=cases, update=yes, warnings=no;
estimate !method=montecarlo, nodes=2000, converge=.01;
export parameters>> D:\10fit-3c-wit(some8.1)52.prm.prm;
export reg_coefficients>> D:\10fit-3c-wit(some8.1)52.reg.reg;
export covariance>> D:\10fit-3c-wit(some8.1)52.cov.cov;
reset;
datafile D:\#52sam10fit-3c.dat;
Format id 1-8 response 11-47;
labels <<D:\#50onet.lab;
codes 0,1,2,3,4;
recode (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
score (0,1,2,3,4) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) !item(2,5,6,25,26,37);
score (0,1,2,3,4) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(3,4,27);
score (0,1,2,3,4) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) !item(1,33);
score (0,1,2,3,4) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) !item(8,9,12,16);
score (0,1,2,3,4) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(10,11,13,15);
score (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) !item(14,17,18,34);
score (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) ( ) (0,1,2,3,4) !item(7);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) (0,1,2,3,4) ( ) ( ) !item(19,20,21,22,35);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) ( ) !item(23,24,30);
score (0,1,2,3,4) ( ) ( ) ( ) ( ) ( ) ( ) (0,1,2,3,4) !item(28,29,31,32,36);
model item+item*step;
set constraints=cases, update=yes, warnings=no;
estimate !method=montecarlo, nodes=2000, converge=.01;
import init_parameters<< D:\10fit-3c-wit(some8.1)52.prm.prm;
import init_reg_coefficients<< D:\10fit-3c-wit(some8.1)52.reg.reg;
import init_covariance<< D:\10fit-3c-wit(some8.1)52.cov.cov;
export parameters>> D:\10fit-3c-wit(some8.1)52.prm.prm;
export reg_coefficients>> D:\10fit-3c-wit(some8.1)52.reg.reg;
export covariance>> D:\10fit-3c-wit(some8.1)52.cov.cov;
show !table=1:2:3:4:5, estimates=latent>> D:\10fit-3c-wit(some8.1)52.shw.shw;
show cases !estimates=eap >> D:\10fit-3c-wit(some8.1)52.eap.eap;
itanal >> D:\10fit-3c-wit(some8.1)52.itn.itn;
```

D1.3 การตรวจสอบความตรงข้ามกลุ่ม

```

Title 10%cross-validation+polytomous+within(some8.1)+onet52;
datafile D:\#52sam10cv-3c.dat;
Format id 1-8 response 11-47;
labels <<D:\#50onet.lab;
codes 0,1,2,3,4;
recode (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
score (0,1,2,3,4) (0,1,2,3,4) () () () () () () !item(2,5,6,25,26,37);
score (0,1,2,3,4) (0,1,2,3,4) () () () () () () (0,1,2,3,4) !item(3,4,27);
score (0,1,2,3,4) () (0,1,2,3,4) () () () () () () !item(1,33);
score (0,1,2,3,4) () () (0,1,2,3,4) () () () () () !item(8,9,12,16);
score (0,1,2,3,4) () () (0,1,2,3,4) () () () () (0,1,2,3,4) !item(10,11,13,15);
score (0,1,2,3,4) () () () (0,1,2,3,4) () () () () !item(14,17,18,34);
score (0,1,2,3,4) () () () (0,1,2,3,4) () () () (0,1,2,3,4) !item(7);
score (0,1,2,3,4) () () () () (0,1,2,3,4) () () () !item(19,20,21,22,35);
score (0,1,2,3,4) () () () () () (0,1,2,3,4) () () !item(23,24,30);
score (0,1,2,3,4) () () () () () () (0,1,2,3,4) () !item(28,29,31,32,36);
model item+item*step;
set constraints=cases, update=yes, warnings=no;
estimate !method=montecarlo, nodes=2000, converge=.01;
import init_parameters << D:\10fit-3c-wit(some8.1)52.prm.prm;
import init_reg_coefficients << D:\10fit-3c-wit(some8.1)52.reg.reg;
import init_covariance << D:\10fit-3c-wit(some8.1)52.cov.cov;
export parameters>> D:\10cv-3c-wit(some8.1)52.prm.prm;
export reg_coefficients>> D:\10cv-3c-wit(some8.1)52.reg.reg;
export covariance>> D:\10cv-3c-wit(some8.1)52.cov.cov;
reset;
datafile D:\#52sam10cv-3c.dat;
Format id 1-8 response 11-47;
labels <<D:\#50onet.lab;
codes 0,1,2,3,4;
recode (0,1,2,3,4) (0,1,2,3,4) !items (1-37);
score (0,1,2,3,4) (0,1,2,3,4) () () () () () () () !item(2,5,6,25,26,37);
score (0,1,2,3,4) (0,1,2,3,4) () () () () () () (0,1,2,3,4) !item(3,4,27);
score (0,1,2,3,4) () (0,1,2,3,4) () () () () () () !item(1,33);
score (0,1,2,3,4) () () (0,1,2,3,4) () () () () () !item(8,9,12,16);
score (0,1,2,3,4) () () (0,1,2,3,4) () () () () (0,1,2,3,4) !item(10,11,13,15);
score (0,1,2,3,4) () () () (0,1,2,3,4) () () () () !item(14,17,18,34);
score (0,1,2,3,4) () () () (0,1,2,3,4) () () () (0,1,2,3,4) !item(7);
score (0,1,2,3,4) () () () () (0,1,2,3,4) () () () !item(19,20,21,22,35);
score (0,1,2,3,4) () () () () () (0,1,2,3,4) () () !item(23,24,30);
score (0,1,2,3,4) () () () () () () (0,1,2,3,4) () !item(28,29,31,32,36);
model item+item*step;
set constraints=cases, update=yes, warnings=no;
estimate !method=montecarlo, nodes=2000, converge=.01;
import init_parameters<< D:\10cv-3c-wit(some8.1)52.prm.prm;
import init_reg_coefficients<< D:\10cv-3c-wit(some8.1)52.reg.reg;
import init_covariance<< D:\10cv-3c-wit(some8.1)52.cov.cov;
show !table=1:2:3:4:5, estimates=latent>> D:\10cv-3c-wit(some8.1)52.shw.shw;
show cases !estimates=eap >> D:\10cv-3c-wit(some8.1)52.eap.eap;
itanal >> D:\10cv-3c-wit(some8.1)52.itn.itn;

```

ตารางที่ D1 รายชื่อผู้เชี่ยวชาญพิจารณาข้อสอบ ONET ปีการศึกษา 2550-2552

1. อาจารย์พรรณวรท วงศาโรจน์	โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ
2. อาจารย์ ดร.ชัยรัตน์ ถาวรงามยิ่งสกุล	โรงเรียนบ้านหนองนกปีกกา จังหวัดตาก
3. อาจารย์สุนทรี นิติเสถียร	โรงเรียนบางระกำวิทยศึกษ
4. อาจารย์วิไลวรรณ อุทธรี	โรงเรียนท่าทองพิทยาคม
5. อาจารย์สุกาย อมรประภา	โรงเรียนพุทธชินราชพิทยา
6. อาจารย์เนาวรัตน์ ประสาร	โรงเรียนเฉลิมขวัญสตรี
7. อาจารย์อัคเคช เผือกตรม	โรงเรียนพุทธชินราชพิทยา
8. อาจารย์สำราญ สุขชี	โรงเรียนจ่านกร้อง
9. อาจารย์จิตราภรณ์ ใจมูลมา	โรงเรียนบางระกำวิทยศึกษ
10. อาจารย์พัชรินทร์ รุ่งท้วม	โรงเรียนท่าทองพิทยาคม
11. อาจารย์พัชรินทร์ สมฤทัย	โรงเรียนพิษณุโลกพิทยาคม
12. อาจารย์นงคราญ เอี่ยมบาง	โรงเรียนพุทธชินราชพิทยา
13. อาจารย์วิไลลักษณ์ งานสดง	โรงเรียนพุทธชินราชพิทยา
14. อาจารย์วารุณี รัตนะ	โรงเรียนเฉลิมขวัญสตรี
15. อาจารย์ชลัยรัตน์ อ๋ามาก	โรงเรียนพิษณุโลกพิทยาคม
16. อาจารย์พรจันทร์เพ็ญ เจ็ญจำรัส	โรงเรียนจุฬาภรณราชวิทยาลัย
17. อาจารย์สมหมาย ปานรูป	โรงเรียนอุตรดิตถ์ดรุณี
18. อาจารย์ฐิติมา พุฒโต	โรงเรียนอุตรดิตถ์ดรุณี
19. อาจารย์เกสินีย์ กรุณานำ	โรงเรียนอุตรดิตถ์
20. อาจารย์กัญปณต จันท์ธินะกุล	โรงเรียนอุตรดิตถ์
21. อาจารย์เกริกฤทธิ์ แดงกองโค	โรงเรียนอุตรดิตถ์
22. อาจารย์ชัยชาญ พรพญาญ	โรงเรียนอุตรดิตถ์
23. อาจารย์อิทธิพงษ์ ตั้งสกุลเรืองไ	สำนักงานพื้นที่การศึกษาเขตมัธยม 39
24. อาจารย์จักรกริสน์ นวลอยู่	โรงเรียนบ้านหนองกุลา
25. อาจารย์เฉลิมพร เทียนศรี	โรงเรียนพิษณุโลกพิทยาคม

ประวัตินักวิจัย

วีรภัทร์ สุขศิริ

กำลังศึกษาในระดับปริญญาเอก สาขาวิชา Quantitative Methods and Evaluation ที่ Graduate School of Education, University of California-Berkeley ประเทศสหรัฐอเมริกา สำเร็จการศึกษาระดับปริญญาโทจาก Lynch School of Education, Boston College สาขา Educational Research, Measurement and Evaluation ประเทศสหรัฐอเมริกา และปริญญาโท ครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลทางการศึกษา จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มีความสนใจเกี่ยวกับ Summative, Formative, และ Diagnostics assessment designs, Test-item development, Personality assessment, Classical test theory และ Item response theory, Advanced Rasch measurement models, และ Generalized linear mixed modeling อีเมลล์: w.suksiri@berkeley.edu

ดร. ชนม์ชกรณ วรอินทร์

สำเร็จการศึกษาดุษฎีบัณฑิต (การทดสอบและวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ ปัจจุบันเป็นอาจารย์ประจำคณะครุศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม จังหวัดพิษณุโลก ปฏิบัติงานตำแหน่งประธานหลักสูตรสาขาวิชาการวิจัยและพัฒนาทางการศึกษา มีประสบการณ์และความชำนาญในด้าน 1) การวิจัยและพัฒนาทางการศึกษา ทางพฤติกรรมศาสตร์และสังคมศาสตร์ 2) การวัดและประเมินผลทางการศึกษา 3) สถิติ การวิเคราะห์ข้อมูลทางการศึกษา ทางพฤติกรรมศาสตร์และสังคมศาสตร์ และ 4) การบริหารจัดการคุณภาพ การประกันคุณภาพองค์กร/หน่วยงาน เบอร์โทรศัพท์ 0-5523-0597 มือถือ 08-5049-4641 E-mail chonchakorn@hotmail.com