

คุณภาพข้อสอบระดับชาติ (โอเน็ต) และแนวทางการพัฒนาคุณภาพ การทดสอบ*

สังวรณั จัตกะโทก**

กาญจนา วัฒนสุนทร***

นลินี ณ นคร****

ศศิธร ชูตินันทกุล*****

ปรารธนา พลอภิชาติ*****

อนุสรณ์ เกิดศรี*****

ปิยนัญ สิทธิฤทธิ์*****

ณภัทร ชัยมงคล*****

นพรัตน์ ไบยา*****

Received: June 16, 2017

Revised: April 27, 2018

Accepted: October 16, 2018

* การวิจัยครั้งนี้ได้รับการสนับสนุนทุนวิจัยจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ประจำปี พ.ศ. 2559

** ค.บ. (มัธยมศึกษา (เคมี-วิทยาศาสตร์)), ค.ม. (การวัดและประเมินผลการศึกษา), Ph.D. (Measurement and Quantitative Methods) Michigan State University, U.S.A. ผู้ช่วยศาสตราจารย์ สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช

*** กศ.บ. (เกียรตินิยม), ค.ม., M.Sc., ค.ด. (การวัดและประเมินผล) จุฬาลงกรณ์มหาวิทยาลัย รองศาสตราจารย์ สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช

**** วท.บ. (ศึกษาศาสตร์) วิชาเอกวัดผลการศึกษา, กศ.ม. (การวัดผลการศึกษา), ค.ด. (การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย ผู้ช่วยศาสตราจารย์ สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช

***** วท.บ. (ศึกษาศาสตร์) วิชาเอกคณิตศาสตร์, กศ.ม. (การวัดผลการศึกษา), ปร.ด. (วิจัย วัดผล และสถิติการศึกษา) มหาวิทยาลัยบูรพา ผู้ช่วยศาสตราจารย์ สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

***** วท.บ. (คณิตศาสตร์), ค.ม. (วิจัยการศึกษา), ค.ด. (การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย อาจารย์ประจำสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

***** วท.บ. (วิทยาการคอมพิวเตอร์), ค.ม. (วิจัยการศึกษา), ค.ด. (การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย อาจารย์ประจำสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

***** กศ.บ. (ภาษาไทย), กศ.ม. (การวิจัยและสถิติทางการศึกษา) อาจารย์ประจำสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

***** วท.บ. (สถิติประยุกต์), วท.ม. (ระบบสารสนเทศเพื่อการจัดการ), ค.ด. (การวัดและประเมินผลการศึกษา) จุฬาลงกรณ์มหาวิทยาลัย ผู้ช่วยผู้อำนวยการ สำนักยุทธศาสตร์และสื่อสารองค์กร

***** ค.บ. (ชีววิทยา), ศษ.ม. (การประเมินการศึกษา) ชำนาญการพิเศษ โรงเรียนบัว จังหวัดน่าน

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) วิเคราะห์คุณภาพข้อสอบระดับชาติขั้นพื้นฐานวิชาภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ระหว่างปีการศึกษา 2554 ถึงปีการศึกษา 2558 และ 2) จัดทำแนวทางการพัฒนาคุณภาพข้อสอบ ข้อมูลที่ใช้คือ ผลการตอบข้อสอบโอเน็ตของนักเรียนชั้นประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 การวิเคราะห์ข้อมูล การวิเคราะห์คุณภาพข้อสอบแบ่งออกเป็น การวิเคราะห์ด้วยแนวคิดการทดสอบแบบดั้งเดิม และการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบ ส่วนการจัดทำแนวทางการพัฒนาคุณภาพการทดสอบดำเนินการโดยวิเคราะห์ข้อมูลเชิงคุณภาพ

ผลการวิจัยพบว่า 1) ผลการวิเคราะห์ข้อสอบด้วยโมเดลการวัดแบบดั้งเดิมและการวิเคราะห์ด้วยโมเดลการตอบข้อสอบ ข้อสอบโอเน็ตส่วนมากมีค่าสถิติผ่านเกณฑ์ทั้งด้านค่าความยาก อำนาจจำแนก ผ่านเกณฑ์ โดยเฉพาะค่าการเดาจากการวิเคราะห์ด้วยโมเดลการตอบข้อสอบนั้นมีค่าค่อนข้างต่ำ ส่วนค่าความเที่ยงนั้นส่วนใหญ่อยู่ระหว่าง 0.6-0.8 ซึ่งถือว่า อยู่ในระดับที่ยอมรับได้ แต่บางปีมีค่าความเที่ยงต่ำ เช่น ปี 2557 วิชาภาษาไทย ระดับชั้นประถมศึกษาปีที่ 6 มีความเที่ยงเท่ากับ .51 ส่วนระดับมัธยมศึกษาปีที่ 3 ที่มีค่าความเที่ยง .37 ซึ่งเป็นแบบทดสอบที่มีจำนวนข้อน้อย ดังนั้น จึงอาจสรุปได้ว่า ข้อสอบโอเน็ตเป็นข้อสอบที่ค่อนข้างมีคุณภาพ สำหรับแนวทางการพัฒนาการทดสอบให้มีคุณภาพดีขึ้นนั้นควรเพิ่มจำนวนข้อให้มากขึ้น และ 2) ในการจัดทำแนวทางการพัฒนาคุณภาพข้อสอบควรกำหนดโครงสร้างข้อสอบให้ครอบคลุมมาตรฐานหลักสูตร และพัฒนาข้อสอบให้สอดคล้องกับโครงสร้างข้อสอบที่กำหนดขึ้น ควรพัฒนาแบบทดสอบแต่ละปีให้เป็นแบบทดสอบคู่ขนานก่อนการนำข้อสอบไปใช้ควรมีการตรวจสอบว่า ข้อสอบวัดได้ตรงตามมาตรฐานหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน โดยนำวิธีการตรวจสอบการวัดความสอดคล้องกับมาตรฐานมาใช้ และควรมีการปรับเทียบคะแนน

คำสำคัญ : ข้อสอบระดับชาติ คุณภาพการทดสอบ การวิเคราะห์ข้อสอบ

ผู้รับผิดชอบบทความ : ผศ.ดร.นลินี ณ นคร สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช 9/9 หมู่ที่ 9 ตำบลบางพูด อำเภอปากเกร็ด จังหวัดนนทบุรี 11120 โทรศัพท์ 0 2504 8573 E-mail : nalinee_na_nakorn@hotmail.com

Analysis of the Ordinary National Educational Test Items (O-NET) and Guidelines on the Development of Test Quality

Sungworn Ngudgratoke*

Kanjana Watanasuntorn**

Nalinee Na Nakorn***

Sasithorn Chutinuntakul****

Prathana Phonapichat*****

Anusorn Koedsri*****

Piyanart Sittirit*****

Nhabhat Chaimongkol*****

Nopparat Baiya*****

* Ph.D. (Measurement and Quantitative Methods), Assistant professor School of Educational Studies, Sukhothai Thammathirat Open University

** Ph.D. (Educational Measurement and Evaluation), Associate professor School of Educational Studies, Sukhothai Thammathirat Open University

*** Ph.D. (Educational Measurement and Evaluation), Assistant professor School of Educational Studies, Sukhothai

**** Ph.D. (Educational Research, Measurement and Statistic), Assistant professor Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

***** Ph.D. (Educational Measurement and Evaluation), Lecture Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

***** Ph.D. (Educational Measurement and Evaluation), Lecture Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

***** M.Ed. (Educational Research and Statistics), Lecture Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

***** Ph.D. (Educational Measurement and Evaluation), Assistant Director Office of Strategy and Organizational Communication

***** Ph.D. Candidate (Educational Evaluation), Teacher Pua School, Nan

Abstract

The objectives of this research were : 1) to analyze the quality of the Ordinary National Education Test (O-NET) items in Thai language subject, Social studies, Religion and Culture subject, English language subject, Mathematics subject, and Sciences subject administered from academic year 2011 (B.E. 2554) to 2015 (B.E. 2558) ; and 2) to provide a guideline on the development of item quality. Data were the O-NET item response from students of Grade 6 (Prathomsuksa 6), Grade 9 (Matthayomsuksa 3) and Grade 12 (Matthayomsuksa 6). Data were analyzed through the classical test theory and the item response theory. Qualitative data analysis was used to produce the guideline on the development of item quality. This research revealed that : 1) the majority of O-NET items, analyzed through the classical test theory and the item response theory, met the difficulty criteria and discrimination criteria. Guessing values from the item response theory were quite low. The reliability coefficients ranged from 0.6 to 0.8 and were acceptable. Yet, there were still low reliability coefficients in certain cases. The Thai language subject in 2014 (B.E. 2557) for Prathomsuksa 6 (Grade 6) had the reliability coefficient of 0.51 while for Matthayomsuksa 3 (Grade 9) it had the reliability coefficients of 0.37. The tests did not have large amount of items. Therefore, based on this finding, it was concluded that O-NET items had fair quality. To develop item quality, it is needs to increase test length; and 2) according to the proposed roadmap on the development of item quality, there should be determining the item structure that covers standards in the curriculum, developing the items closely relevant to the determined item structure, developing the parallel tests in each year, examining the alignment between the items and the standards in the Basic Education Core Curriculum by using the alignment evaluation method, and equating the test scores across years.

Keywords : Ordinary National Educational Test, Test quality, Item analysis

Corresponding Author : Asst. Prof. Nalinee Na Nakorn, School of Educational Studies, Sukhothai Thammathirat Open University, 9/9 Moo 9 Chaengwattana Rd. Bangpood, Pakkret, Nonthaburi 11120. Tel 0 2504 8573 E-mail : nalinee_na_nakorn@hotmail.com

บทนำ

สถาบัน ทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ได้รับการจัดตั้งขึ้นตามพระราชกฤษฎีกา เมื่อ พ.ศ. 2548 เพื่อบริหารจัดการเกี่ยวกับการทดสอบความรู้ทางการศึกษา ระดับชาติของนักเรียนระดับการศึกษาขั้นพื้นฐาน มีพันธกิจหลักคือ การทดสอบและประเมินผลทางการศึกษาระดับชาติ โดยได้จัดการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (Ordinary National Educational Test : O-NET) หรือ “โอเน็ต” ตั้งแต่ พ.ศ. 2548 ซึ่งเป็นการวัดและประเมินผลการเรียนรู้รวมยอดของผู้เรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน การดำเนินงานของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ไม่ได้เน้นเพียงการดำเนินการทดสอบเท่านั้น แต่ สทศ. ยังสะท้อนการดำเนินงานโดยการนำผลการทดสอบไปใช้ชี้แนะผู้รับบริการ (customers) นักเรียนและผู้ปกครอง (users) ให้รับรู้ว่า นักเรียนได้อะไร และผู้ปกครองได้อะไรจากผลการทดสอบ ซึ่ง สทศ. ได้พยายามสร้างรูปแบบการรายงานผลการประเมินการศึกษาระดับชาติเพื่อให้สถานศึกษาและหน่วยงานที่เกี่ยวข้องได้รับรู้ผลการประเมินที่สามารถนำไปใช้ในการดำเนินงานพัฒนาคุณภาพการศึกษาของประเทศ

ปัจจุบัน สทศ. มีแนวคิดเกี่ยวกับการพัฒนาระบบการรายงานผลการประเมินให้ผู้เกี่ยวข้องได้รับรู้อย่างชัดเจน และส่งเสริมการนำผลการประเมินไปใช้มากขึ้น เพื่อให้การดำเนินการในทิศทางนี้ประสบความสำเร็จ สทศ. มีความประสงค์ให้มีการดำเนินงานจัดทำรายงานผลการทดสอบระดับชาติ เพื่อได้ข้อเสนอแนะเชิงนโยบายกับหน่วยงานที่เกี่ยวข้องทั้งในระดับมหภาค ได้แก่ รัฐบาล กระทรวงศึกษาธิการ และหน่วยงานที่รับผิดชอบการจัดการ

ศึกษา ได้แก่ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และสำนักงานเขตพื้นที่การศึกษา เพื่อนำไปสู่การพัฒนาการเรียนการสอนระดับโรงเรียนด้วยเหตุที่ สทศ. มีความจำเป็นต้องดำเนินการจัดทำโครงการดังกล่าว การวิจัยเรื่องนี้จึงต้องการตอบสนองการดำเนินงานของ สทศ. โดยนำข้อมูลผลการสอบโอเน็ตย้อนหลัง 5 ปี มาวิเคราะห์ข้อมูลทางสถิติทั้งในด้านคุณภาพข้อสอบ และการวินิจฉัยคุณภาพการเรียนของผู้เรียนและสถานศึกษา เพื่อให้ได้ข้อมูลสำหรับเป็นแนวทางกำหนดนโยบายส่งเสริมการรายงานผลการทดสอบระดับชาติแก่สถานศึกษา และหน่วยงานที่เกี่ยวข้องอย่างมีประสิทธิภาพและสถานศึกษาได้พัฒนาผู้เรียนให้มีคุณภาพตามหลักสูตรต่อไป

วัตถุประสงค์

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ

- 1) วิเคราะห์คุณภาพข้อสอบระดับชาติ (โอเน็ต) วิชาภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ตั้งแต่ปีการศึกษา พ.ศ. 2554 ถึง 2558
- 2) จัดทำแนวทางการพัฒนาคุณภาพการทดสอบโอเน็ต

กรอบแนวคิดของการวิจัย

กรอบแนวคิดของการวิจัยในครั้งนี้มี 2 ประการคือ เป็นการวิเคราะห์คุณภาพข้อสอบ โดยมีการนำผลการสอบโอเน็ตย้อนหลัง 5 ปี มาวิเคราะห์เพื่อให้ทราบถึงคุณภาพของการพัฒนาข้อสอบ และข้อเสนอแนะในการพัฒนาข้อสอบโอเน็ตให้มีคุณภาพมากขึ้นในอนาคต การวิเคราะห์ข้อสอบจะมีการวิเคราะห์ทั้งทฤษฎี

การทดสอบแบบดั้งเดิม และทฤษฎีการตอบข้อสอบ เพื่อให้ได้สารสนเทศของข้อสอบที่ชัดเจนสำหรับเป็นข้อมูลในการพัฒนาการสร้างข้อสอบต่อไป

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้นำข้อมูลทฤษฎีการสอบโอเน็ต มาวิเคราะห์เพื่อนำเสนอนโยบายการรายงานผลการทดสอบระดับชาติแก่สถานศึกษาเพื่อพัฒนาผู้เรียน ในการอธิบายวิธีดำเนินการวิจัย ผู้วิจัยอธิบายวิธีดำเนินการวิจัยแยกตามวัตถุประสงค์ของการวิจัย ดังต่อไปนี้

วัตถุประสงค์ที่ 1 เพื่อวิเคราะห์คุณภาพข้อสอบระดับชาติขั้นพื้นฐานวิชาภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ตั้งแต่ปีการศึกษา พ.ศ. 2554 ถึง 2558

ข้อมูลที่ใช้ ในการวิเคราะห์คุณภาพข้อสอบระดับชาติขั้นพื้นฐานย้อนหลัง 5 ปีคือ ผลการสอบโอเน็ตรายข้อวิชา จำนวน 5 ชุดวิชาคือ วิชาภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษา

อังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ตั้งแต่ปีการศึกษา พ.ศ. 2554 ถึง 2558

การวิเคราะห์ข้อมูล การวิเคราะห์คุณภาพข้อสอบแบ่งออกเป็น การวิเคราะห์ด้วยแนวคิดการทดสอบแบบดั้งเดิม (Classical Test Theory) และการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบ (Item Response Theory) โดยการวิเคราะห์ด้วยแนวคิดการทดสอบแบบดั้งเดิมเป็นการรายงานค่าความยาก อำนาจจำแนก (พอยต์ไบซีเรียล) ความเที่ยง (reliability) ส่วนการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบเป็นการรายงานค่าสถิติข้อสอบ (ความยาก อำนาจจำแนก และค่าการเดา)

การวิเคราะห์ข้อสอบด้วยแนวคิดการทดสอบแบบดั้งเดิมใช้ตัวอย่างทั้งหมด แต่เนื่องจากการวิเคราะห์ด้วยทฤษฎีการตอบข้อสอบนั้นมีข้อจำกัดในการวิเคราะห์ เมื่อมีผู้สอบจำนวนมากจึงใช้การสุ่มผู้สอบมาจำนวนจังหวัดละ 154 คน ตามแนวทางของ The Programme for International Student Assessment : PISA (OECD, 2012) รวม 11,704 คน โดยเกณฑ์การประเมินคุณภาพข้อสอบมีดังนี้

1. เกณฑ์การแปลผลค่าความยาก (p)

เกณฑ์	การแปลความหมาย
0.75-1.0	ง่าย
0.25-0.74	ใช้ได้
ต่ำกว่า 0.25	ยาก

2. เกณฑ์การแปลผลค่าอำนาจจำแนกของข้อสอบ (พอยต์ไบซีเรียล) ใช้เกณฑ์ของ Ebel และ Frisbie (1986)

เกณฑ์	พอยต์ไบซีเรียล	การแปลความหมาย
ดีมาก	มากกว่า 0.39	ใช้ได้
ดี	0.30-0.39	อาจปรับแก้ให้ดีขึ้น
ปานกลาง	0.20-0.29	ควรทบทวน
ต่ำ	0.00-0.20	ควรเอาทิ้ง หรือทบทวนอย่างละเอียด
ใช้ไม่ได้	ต่ำกว่า -0.01	ควรเอาทิ้ง

3. เกณฑ์การแปลผลค่าความยาก (difficulty) ใช้เกณฑ์ของ Baker & Kim (2004)

ค่า	การแปลผล
+1.00 เป็นต้นไป	ยาก หรือเหมาะกับผู้สอบที่เก่ง
-0.99-0.99	ปานกลาง หรือเหมาะกับผู้สอบที่เก่งปานกลาง
-1.00 หรือต่ำกว่า	ง่าย หรือเหมาะกับผู้สอบที่ไม่เก่ง

4. เกณฑ์การแปลผลค่าอำนาจจำแนก (discrimination) ใช้เกณฑ์ของ Baker & Kim (2004)

ค่า	การแปลผล
0.01-0.34	ต่ำมาก
0.35-0.64	ต่ำ
0.65-1.34	ปานกลาง
1.35-1.69	สูง
1.70 ขึ้นไป	สูงมาก

5. เกณฑ์การแปลผลค่าการเดา (guessing)

ค่า	การแปลผล
0.0-0.1	ต่ำ
0.10-0.20	ปานกลาง
มากกว่า 0.20	สูง

วัตถุประสงค์ที่ 2 การจัดทำแนวทางการพัฒนาคุณภาพการทดสอบโอเน็ต

ข้อมูลที่ต้องการ ผลการวิเคราะห์คุณภาพข้อสอบรายข้อวิชาภาษาไทย สังคมศึกษา ศาสนา และวัฒนธรรม ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ตั้งแต่ปีการศึกษา พ.ศ. 2554 ถึง 2558 และข้อมูลข้อเสนอแนะจากผู้มีส่วนได้-เสียจำนวน 50 คน ประกอบด้วย ครู ผู้บริหารสถานศึกษา เขตพื้นที่การศึกษา นักวิชาการ และผู้ปกครอง

การวิเคราะห์ข้อมูล วิเคราะห์ข้อมูลด้วยการสังเคราะห์ผลการวิเคราะห์ข้อสอบ การวิเคราะห์เนื้อหาและสังเคราะห์ข้อมูลเกี่ยวกับการพัฒนาคุณภาพข้อสอบตามแนวคิดของการประเมินอิงมาตรฐาน (standards-based assessment) แนวคิดการประกันคุณภาพการพัฒนาข้อสอบ แนวคิดการประเมินทางการศึกษาตามนโยบาย common core standards ของสหรัฐอเมริกา มาตรฐานของการทดสอบทางการศึกษาและจิตวิทยา และมาตรฐานการทดสอบระดับนานาชาติ และวิเคราะห์ข้อมูลเชิงคุณภาพจากผู้มีส่วนได้-เสียด้วยการวิเคราะห์เนื้อหา

ผลการวิจัย

ผู้วิจัยสรุปผลการวิจัยตามวัตถุประสงค์การวิจัยดังนี้

1. เพื่อวิเคราะห์คุณภาพข้อสอบระดับชาติขั้นพื้นฐานวิชาภาษาไทย สังคมศึกษา ศาสนา และวัฒนธรรม ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ตั้งแต่ปีการศึกษา พ.ศ. 2554 ถึง 2558

ผลการวิเคราะห์ข้อสอบด้วยโมเดลการวัดแบบดั้งเดิม และการวิเคราะห์ด้วยโมเดลการตอบ

ข้อสอบพบว่า ข้อสอบโอเน็ตส่วนมากมีค่าสถิติผ่านเกณฑ์ทั้งด้านค่าความยาก อำนาจจำแนก ผ่านเกณฑ์ โดยเฉพาะค่าการเดาจากการวิเคราะห์ด้วยโมเดล IRT นั้นมีค่าค่อนข้างต่ำ หรือนักเรียนเดาไม่ค่อยถูก ส่วนค่าความเที่ยงนั้นส่วนใหญ่อยู่ระหว่าง 0.6-0.8 ซึ่งถือว่า อยู่ในระดับที่ยอมรับได้ แต่พบว่าบางปีมีค่าความเที่ยงต่ำ เช่น วิชาภาษาไทย ระดับมัธยมศึกษาปีที่ 3 ปี 2557 ที่มีค่าความเที่ยง 0.37 ส่วนระดับชั้นประถมศึกษาปีที่ 6 มีความเที่ยงเท่ากับ 0.51 ผลการวิเคราะห์ดังตารางที่ 1 - 2

2. จัดทำแนวทางการพัฒนาคุณภาพการทดสอบโอเน็ต

แนวทางการพัฒนาข้อสอบที่สำคัญคือ 1) การพัฒนาข้อสอบให้สามารถวัดได้ตรงตามมาตรฐานหลักสูตรและตรงตามโครงสร้างข้อสอบที่กำหนดขึ้น และควรพัฒนาข้อสอบให้เป็นข้อสอบคู่ขนาน ก่อนการนำข้อสอบไปใช้ควรมีการตรวจสอบว่า ข้อสอบวัดได้ตรงตามมาตรฐานหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานหรือไม่ โดยนำวิธีการตรวจสอบการวัดความสอดคล้องกับมาตรฐาน (alignment) มาใช้ และ 2) การปรับเทียบคะแนนสอบระหว่างปีเพื่อให้เทียบเคียงกันได้ เพื่อการเปรียบเทียบคุณภาพการจัดการศึกษาของโรงเรียนมีความเหมาะสมและถูกต้องมากขึ้น โดยอาจกำหนดให้ปี 2558 เป็นปีฐาน และปรับเทียบคะแนนของปีต่อไปให้เทียบเคียงกับคะแนนสอบที่เป็นปีฐาน (2558) ทั้งนี้การปรับเทียบคะแนนควรมีการออกแบบการทดสอบให้มีข้อสอบร่วมในการสอบระหว่างปีโดยไม่เอาคะแนนข้อสอบร่วมมาคิดคะแนนรวม หรือเป็นการใช้ข้อสอบร่วมภายนอก (external anchor items)

ทั้งนี้ ตัวแทนสถานศึกษา (ครู และผู้อำนวยการสถานศึกษา) ผู้ปกครอง และเขตพื้นที่การศึกษาจังหวัดต่าง ๆ ได้ให้ข้อเสนอแนะการใช้ผลการทดสอบระดับชาติเพื่อพัฒนาผู้เรียนให้มีคุณภาพ ดังนี้

1) ครู โรงเรียน และเขตพื้นที่การศึกษา ต้องการให้มีการรายงานผลการทดสอบของนักเรียน เป็นรายบุคคลอย่างละเอียดถึงคะแนนรายตัวชี้วัด เพื่อจะได้ทำการวิเคราะห์ความสามารถของนักเรียน

ได้ละเอียดมากขึ้น ส่วนผู้ปกครองต้องการการรายงานผลที่ไม่ซับซ้อน เข้าใจง่าย

2) ผู้ที่เกี่ยวข้องเสนอแนะให้ สทศ. ดำเนินการพัฒนาข้อสอบให้มีคุณภาพ ทั้งในด้านการกำหนดโครงสร้างข้อสอบให้ชัดเจน การใช้โครงสร้างเดียวกันในการออกข้อสอบแต่ละปีเพื่อให้เทียบเคียงผลการสอบระหว่างปีได้ และเพื่อให้การเขียนข้อสอบ ตรงตามโครงสร้างที่กำหนด

ตารางที่ 1

ผลการวิเคราะห์ข้อสอบด้วยโมเดลการวัดแบบดั้งเดิม

วิชา	ประถมศึกษาปีที่ 6			มัธยมศึกษาปีที่ 3			มัธยมศึกษาปีที่ 6		
	ความยาก	อำนาจจำแนก	ความเที่ยง	ความยาก	อำนาจจำแนก	ความเที่ยง	ความยาก	อำนาจจำแนก	ความเที่ยง
ปีการศึกษา 2554									
ภาษาไทย	0.22 – 0.79	0.04 – 0.39	0.86	0.03 – 0.88	-0.09 – 0.48	0.78	0.03 – 0.92	-0.20 – 0.46	0.83
สังคมศึกษา ศาสนาและวัฒนธรรม	0.18 – 0.87	-0.08 – 0.54	0.88	0.19 – 0.85	-0.08 – 0.35	0.67	0.06 – 0.78	-0.11 – 0.37	0.66
ภาษาอังกฤษ	0.24 – 0.62	0.13 – 0.41	0.73	0.19 – 0.56	-0.11 – 0.32	0.66	0.09 – 0.52	-0.12 – 0.50	0.86
คณิตศาสตร์	0.17 – 0.86	0.14 – 0.49	0.88	0.01 – 0.60	0.09 – 0.32	0.57	0.02 – 0.59	0.04 – 0.61	0.81
วิทยาศาสตร์	0.10 – 0.71	-0.06 – 0.38	0.86	0.01 – 0.88	0.00 – 0.50	0.70	0.06 – 0.70	-0.05 – 0.48	0.70
ปีการศึกษา 2555									
ภาษาไทย	0.23 – 0.65	0.04 – 0.37	0.63	0.14 – 0.74	-0.10 – 0.41	0.71	0.12 – 0.82	0.04 – 0.50	0.89
สังคมศึกษา ศาสนาและวัฒนธรรม	0.17 – 0.74	0.03 – 0.44	0.69	0.21 – 0.68	-0.04 – 0.47	0.73	0.09 – 0.66	-0.04 – 0.28	0.68
ภาษาอังกฤษ	0.28 – 0.59	0.13 – 0.43	0.62	0.21 – 0.45	0.04 – 0.28	0.69	0.04 – 0.45	-0.05 – 0.45	0.87
คณิตศาสตร์	0.08 – 0.75	0.09 – 0.46	0.74	0.06 – 0.49	0.03 – 0.22	0.48	0.03 – 0.38	0.06 – 0.55	0.80
วิทยาศาสตร์	0.06 – 0.68	0.07 – 0.30	0.83	0.11 – 0.61	0.06 – 0.37	0.69	0.05 – 0.64	-0.01 – 0.43	0.82
ปีการศึกษา 2556									
ภาษาไทย	0.08 – 0.62	0.06 – 0.30	0.63	0.03 – 0.79	-0.14 – 0.38	0.70	0.08 – 0.82	-0.01 – 0.46	0.89
สังคมศึกษา ศาสนาและวัฒนธรรม	0.11 – 0.71	-0.02 – 0.32	0.75	0.21 – 0.69	-0.06 – 0.34	0.66	0.09 – 0.50	-0.02 – 0.38	0.72
ภาษาอังกฤษ	0.20 – 0.50	0.01 – 0.40	0.62	0.23 – 0.45	0.01 – 0.30	0.64	0.04 – 0.58	-0.06 – 0.43	0.88
คณิตศาสตร์	0.25 – 0.70	0.08 – 0.49	0.70	0.01 – 0.53	0.06 – 0.30	0.55	0.06 – 0.36	0.11 – 0.49	0.81
วิทยาศาสตร์	0.15 – 0.58	0.08 – 0.29	0.66	0.13 – 0.57	0.00 – 0.42	0.78	0.04 – 0.54	-0.06 – 0.37	0.77

ตารางที่ 1 (ต่อ)

วิชา	ประเมินศึกษาปีที่ 6		มัธยมศึกษาปีที่ 3		มัธยมศึกษาปีที่ 6	
	ความยาก	อำนาจจำแนก	ความเที่ยง	ความยาก	อำนาจจำแนก	ความเที่ยง
ปีการศึกษา 2557						
ภาษาไทย	0.08 – 0.80	0.03 – 0.29	0.51	0.01 – 0.70	-0.06 – 0.23	0.37
สังคมศึกษา ศาสนาและวัฒนธรรม	0.23 – 0.78	0.08 – 0.46	0.81	0.11 – 0.79	0.00 – 0.34	0.67
ภาษาอังกฤษ	0.27 – 0.49	0.21 – 0.43	0.85	0.16 – 0.45	-0.01 – 0.23	0.59
คณิตศาสตร์	0.11 – 0.67	0.10 – 0.43	0.69	0.06 – 0.60	-0.02 – 0.41	0.64
วิทยาศาสตร์	0.13 – 0.72	0.08 – 0.30	0.67	0.09 – 0.57	0.01 – 0.36	0.74
ปีการศึกษา 2558						
ภาษาไทย	0.13 – 0.83	0.07 – 0.29	0.69	0.05 – 0.71	-0.05 – 0.30	0.56
สังคมศึกษา ศาสนาและวัฒนธรรม	0.26 – 0.73	0.07 – 0.40	0.76	0.20 – 0.78	0.03 – 0.36	0.73
ภาษาอังกฤษ	0.22 – 0.74	0.07 – 0.45	0.86	0.20 – 0.44	0.00 – 0.37	0.71
คณิตศาสตร์	0.19 – 0.73	0.09 – 0.54	0.77	0.05 – 0.60	0.13 – 0.37	0.73
วิทยาศาสตร์	0.13 – 0.68	0.03 – 0.30	0.62	0.13 – 0.57	-0.06 – 0.45	0.75

ตารางที่ 2

ผลการวิเคราะห์ข้อสอบด้วยโมเดลการตอบข้อสอบ

วิชา	ประเมินศึกษาปีที่ 6		มัธยมศึกษาปีที่ 3		มัธยมศึกษาปีที่ 6	
	ความยาก	อำนาจจำแนก	การเดา	ความยาก	อำนาจจำแนก	การเดา
ปีการศึกษา 2554						
ภาษาไทย	-1.82 – 5.00	0.36 – 2.07	0.01 – 0.33	-4.34 – 5.00	0.16 – 4.24	0.01 – 0.33
สังคมศึกษา ศาสนาและวัฒนธรรม	-2.53 – 5.00	0.10 – 2.48	0.01 – 0.35	-5.00 – 5.00	0.15 – 3.02	0.01 – 0.32
ภาษาอังกฤษ	-1.82 – 5.00	0.36 – 2.96	0.01 – 0.43	0.28 – 5.00	0.16 – 4.47	0.10 – 0.34
คณิตศาสตร์	-2.53 – 2.52	0.68 – 2.91	0.01 – 0.35	0.29 – 5.00	0.23 – 4.25	0.01 – 0.46
วิทยาศาสตร์	-1.22 – 5.00	0.47 – 4.94	0.01 – 0.37	-3.61 – 2.45	0.33 – 5.00	0.01 – 0.54

ตารางที่ 2 (ต่อ)

วิชา	ประถมศึกษาปีที่ 6				มัธยมศึกษาปีที่ 3				มัธยมศึกษาปีที่ 6			
	ความยาก	อำนาจจำแนก	การเดา	การเดา	ความยาก	อำนาจจำแนก	การเดา	การเดา	ความยาก	อำนาจจำแนก	การเดา	การเดา
ปีการศึกษา 2555												
ภาษาไทย	-5.00 – 5.00	0.00 – 3.11	0.01 – 0.40		-2.17 – 5.00	0.02 – 5.00	0.01 – 0.34		-2.20 – 5.00	0.07 – 5.00	0.01 – 0.53	
สังคมศึกษา ศาสนาและวัฒนธรรม	-2.48 – 5.00	0.06 – 5.00	0.01 – 0.62		-5.00 – 5.00	0.00 – 5.00	0.01 – 0.39		-1.10 – 5.00	0.00 – 5.00	0.01 – 0.47	
ภาษาอังกฤษ	-0.42 – 2.21	1.15 – 5.00	0.01 – 0.51		0.99 – 5.00	0.34 – 4.32	0.01 – 0.40		-1.30 – 5.00	0.23 – 5.00	0.01 – 0.29	
คณิตศาสตร์	-0.75 – 1.49	0.62 – 5.00	0.01 – 0.54		0.45 – 5.00	0.00 – 5.00	0.01 – 0.31		-0.99 – 0.26	0.39 – 4.82	0.01 – 0.23	
วิทยาศาสตร์	-0.83 – 5.00	0.21 – 3.19	0.01 – 0.35		-0.14 – 5.00	0.10 – 5.00	0.01 – 0.57		-0.74 – 5.00	0.03 – 5.00	0.01 – 0.44	
ปีการศึกษา 2556												
ภาษาไทย	-0.81 – 5.00	0.34 – 1.58	0.01 – 0.28		-1.62 – 5.00	0.12 – 5.00	0.01 – 0.48		-0.02 – 5.00	0.15 – 5.00	0.01 – 0.53	
สังคมศึกษา ศาสนาและวัฒนธรรม	-1.12 – 5.00	0.13 – 4.55	0.01 – 0.39		-1.49 – 5.00	0.01 – 5.00	0.01 – 0.41		-0.46 – 5.00	0.00 – 4.78	0.01 – 0.39	
ภาษาอังกฤษ	0.48 – 3.12	0.67 – 5.00	0.01 – 0.38		0.63 – 3.01	0.30 – 4.01	0.01 – 0.33		-1.48 – 5.00	0.76 – 5.00	0.01 – 0.33	
คณิตศาสตร์	-1.75 – 1.68	0.19 – 3.11	0.01 – 0.25		-0.11 – 5.00	0.26 – 5.00	0.01 – 0.34		-0.33 – 0.85	1.07 – 5.00	0.01 – 0.24	
วิทยาศาสตร์	-0.78 – 5.00	0.10 – 3.48	0.01 – 0.34		-0.12 – 5.00	0.03 – 4.55	0.01 – 0.48		-0.09 – 5.00	0.00 – 5.00	0.01 – 0.52	
ปีการศึกษา 2557												
ภาษาไทย	-5.00 – 5.00	0.00 – 3.29	0.01 – 0.31		-0.16 – 5.00	0.06 – 4.73	0.01 – 0.66		-2.40 – 5.00	0.15 – 2.86	0.01 – 0.37	
สังคมศึกษา ศาสนาและวัฒนธรรม	-2.58 – 5.00	0.10 – 5.00	0.01 – 0.45		-5.00 – 5.00	0.00 – 4.75	0.01 – 0.51		-1.09 – 5.00	0.00 – 5.00	0.01 – 0.51	
ภาษาอังกฤษ	0.55 – 1.89	0.81 – 4.73	0.15 – 0.36		1.40 – 5.00	0.29 – 4.04	0.10 – 0.42		-1.07 – 2.10	0.36 – 5.00	0.01 – 0.32	
คณิตศาสตร์	-0.59 – 2.22	0.55 – 5.00	0.01 – 0.52		-1.29 – 2.06	0.42 – 5.00	0.01 – 0.55		-0.78 – 0.70	1.30 – 5.00	0.01 – 0.24	
วิทยาศาสตร์	-1.28 – 2.84	0.45 – 5.00	0.01 – 0.53		-0.05 – 5.00	0.18 – 4.15	0.01 – 0.33		-3.84 – 5.00	0.07 – 5.00	0.01 – 0.51	
ปีการศึกษา 2558												
ภาษาไทย	-5.00 – 5.00	0.00 – 2.71	0.01 – 0.34		-1.94 – 5.00	0.07 – 5.00	0.01 – 0.25		-1.86 – 2.92	0.22 – 5.00	0.01 – 0.53	
สังคมศึกษา ศาสนาและวัฒนธรรม	-5.00 – 5.00	0.00 – 3.28	0.01 – 0.40		-5.00 – 5.00	0.00 – 5.00	0.01 – 0.46		-5.00 – 5.00	0.00 – 5.00	0.01 – 0.62	
ภาษาอังกฤษ	-0.52 – 3.07	1.24 – 4.52	0.08 – 0.40		0.72 – 5.00	0.27 – 5.00	0.01 – 0.33		-0.44 – 5.00	0.13 – 5.00	0.01 – 0.29	
คณิตศาสตร์	-1.18 – 1.68	0.76 – 4.92	0.01 – 0.42		0.26 – 2.77	1.14 – 4.71	0.01 – 0.46		-1.01 – 5.00	0.66 – 5.00	0.01 – 0.36	
วิทยาศาสตร์	-0.64 – 5.00	0.21 – 4.82	0.01 – 0.62		-0.38 – 5.00	0.25 – 5.00	0.01 – 0.45		-4.97 – 5.00	0.00 – 5.00	0.01 – 0.43	

อภิปรายผลการวิจัย

จากผลการวิจัย ผู้วิจัยมีประเด็นในการอภิปรายผลการวิจัยในประเด็นคุณภาพข้อสอบโอเน็ตและแนวทางการพัฒนาคุณภาพข้อสอบที่สำคัญ ดังนี้

ข้อสอบโอเน็ตที่พัฒนาขึ้นในภาพรวมมีคุณภาพในด้านความยาก อำนาจจำแนก ค่าพารามิเตอร์การเดา ประเด็นที่ยังเป็นปัญหา คือ ความเที่ยงซึ่งเป็นเพราะเนื้อหาของวิชานั้นมีความหลากหลายและแตกต่างกัน การคำนวณค่าความเที่ยงรวมจึงทำให้ค่าความเที่ยงต่ำเกินจริง การแก้ปัญหาคุณภาพด้านความเที่ยงอาจทำได้หลายทาง เช่น 1) รายงานความเที่ยงของแต่ละมาตรฐาน หรือสาระที่ 2) การเพิ่มจำนวนข้อสอบให้มากขึ้น จากผลการวิเคราะห์แบบทดสอบที่มีจำนวนข้อ 50 ข้อขึ้นไป ไม่ค่อยมีปัญหาด้านความเที่ยง ดังนั้น แบบทดสอบโอเน็ตบางวิชาที่มีข้อสอบน้อยควรกำหนดให้เพิ่มข้อสอบมากขึ้นเป็น 50 ข้อ ส่วนการปรับปรุงและพัฒนาข้อสอบให้มีคุณภาพ

ด้านความตรงควรมีการวิเคราะห์ความสอดคล้องของข้อสอบและมาตรฐานก่อนนำไปใช้ เช่น การนำวิธีการตรวจสอบความสอดคล้องของเนื้อหากับมาตรฐานในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานตามวิธีการของ Webb (1997, 1999) มาใช้เพื่อประเมินว่า ข้อสอบมีความสอดคล้องกับมาตรฐานในด้านเนื้อหา ความลึกของเนื้อหา การกระจายครอบคลุมเนื้อหา และความสมดุลการดำเนินการเหล่านี้ใช้เวลาและใช้ผู้ทรงคุณวุฒิมาก ดังนั้น การพัฒนาข้อสอบโอเน็ตควรดำเนินการพัฒนาและตรวจสอบก่อนนำไปใช้ประมาณ 1-2 ปี ถึงแม้เวลาในการพัฒนาจะใช้เวลามาก แต่เป็นการรับประกันคุณภาพข้อสอบว่า ข้อสอบที่พัฒนาขึ้นมีความตรงและสอดคล้องกับมาตรฐาน ซึ่งจะช่วยลดปัญหาคุณภาพข้อสอบที่เกิดขึ้นในปัจจุบันได้ และในการรายงานผลการสอบควรมีการเปรียบเทียบคะแนนระหว่างปีเพื่อให้สามารถเปรียบเทียบคะแนนสอบระหว่างปีได้อย่างเป็นธรรม และถูกต้อง 

References

- Baker, F. B., & Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques* (2nd ed.). New York: Marcel Dekker.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of Educational Measurement*. (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall
- OECD. (2012). *PISA 2012 Technical Report*. Organization for Economic Co-operation and Development.
- Webb, N. L. (1997). *Research Monograph No.8. Criteria for Aligning of Expectations and Assessments in Mathematics and Science Education*. Retrieved June 19th, 2012, from <http://facstaff.wceruw.org/normw/WEBBMonograph6criteria.pdf>.
- _____. (1999). *Research Monograph No.18. Alignment of Science and Mathematics Standards and Assessments in Four States*. Washington, DC: Council of Chief State School Officers.