

THE DEVELOPMENT AND ANALYSIS OF THE QUALITY OF GROWTH MEASUREMENT METHODS BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY*

Auyporn Ruengrakul

ABSTRACT

The purposes of this research were to develop, analyze, and compare the quality of the nine methods measuring growth. The nine growth measurement methods were five methods based on the classical test theory, two methods based on the item response theory, and two methods developed by the researcher. The two samples consisted of 698 and 637 Mathayom Suksa 3 students, studying in the first term of 1999 academic year, in 12 schools under the Department of General Education, Bangkok Metropolis. The research instrument were 3 sets of parallel mathematics tests for Mathayom Suksa 3 students. There were 3 data sets for this research. The first two sets were primary data obtaining from three - time - point measurement of mathematics achievement, with a four - week time - lag. Both two data sets were repeated measures, using the same tests for the first one, and using parallel tests with latin - square design for the second one. The third data set was secondary data obtaining from five - time - point measurement of mathematics achievement at Mathayom Suksa 2 level. All three data sets had been scored employing dichotomous and partial credit scoring approaches, except the third data set employed only dichotomous scoring methods. Therefore, there were altogether 7 data bases for this study. Data were analyzed in order to estimate growth score by PARSCALE 2, BILOG 3, LPCM - WIN 1.0, LISREL 8.3, and to compare the quality among the methods growth measurement by HLM - WIN 5.04, SPSS - WIN 10.07 and EXCEL.

The research results indicated that when comparing the growth score estimations across 3 theoretical approaches, there were no significance differences among the classical test theory approach, the item response theory approach and the researcher's development approach. The comparison among 5 methods within the classical test theory approach revealed that the methods of observed difference score and the relative growth method had better quality than other methods for the first 2 data sets; and that the residual growth and potential ratio method had better quality than other methods for the third data set. The comparison among 2 methods within the item response theory approach revealed that the method of multidimensional Rasch model for learning and change had better quality than other method. The comparison of 2 methods within the researcher's development approach revealed that the relative true ability of growth and initial ability ratio method had better quality than other method for dichotomous data sets.

* Doctoral dissertation of Department of Educational Research, Chulalongkorn University under the advice of Assoc. Prof. Sirichai Kanjanawasee, Ph.D. and Prof. Nonglak Wiratchai, Ph.D.

การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนน พัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎี การตอบสนองข้อสอบ*

อวยพร เรื่องตระกูล

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาและวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการรวม 9 วิธี ประกอบด้วย วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธี กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม กลุ่มละ 698 คน และ 637 คน ในภาคเรียนที่ 1 ปีการศึกษา 2543 จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร 12 โรงเรียน เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่คู่ขนานกัน ข้อมูลที่ใช้ในการวิจัยมี 3 ชุด ชุดแรกเป็นข้อมูลปฐมภูมิ ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ที่มีการเก็บรวบรวมข้อมูลโดยการวัด 3 ครั้ง ข้อมูลชุดแรกเป็นการวัดซ้ำด้วยแบบทดสอบฉบับเดิม ข้อมูลชุดที่สองเป็นการวัดซ้ำแบบสุ่มอิสระด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง แยกวิเคราะห์เป็นแบบวัดซ้ำ 3 และ 5 ครั้ง ข้อมูลทั้ง 3 ชุด มีการตรวจให้คะแนนแบบทวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) ยกเว้นข้อมูลชุดที่สามแบบวัดซ้ำ 5 ครั้งที่มีการตรวจให้คะแนนแบบทวิภาคเท่านั้น ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงมีฐานข้อมูลทั้งหมด 7 ฐาน การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการด้วยโปรแกรม PARSCALE 2, BILOG 3, LPCM-WIN 1.0, LISREL 8.3 และการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการด้วยโปรแกรม HLM - WIN 5.04, SPSS - WIN 10.07 และ EXCEL

ผลการวิจัยสรุปได้ว่า เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีคุณภาพไม่แตกต่างกัน ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่า วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และ วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3 ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์พหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงมีคุณภาพสูงกว่าวิธีอื่น และผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนมีคุณภาพสูงกว่าวิธีอื่น เมื่อมีการตรวจให้คะแนนแบบทวิภาค

* อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาร่วม ศาสตราจารย์กิตติคุณ ดร. นงลักษณ์ วิรัชชัย ส่วนหนึ่งของวิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา ปีการศึกษา

ความเป็นมาและความสำคัญของปัญหา

การวัดคะแนนพัฒนาการเป็นวิธีวิทยาการวัดที่ได้รับความสนใจจากศาสตร์เกือบทุกสาขา จุดมุ่งหมายในการวัดคะแนนพัฒนาการที่สำคัญคือเพื่อพัฒนาวิธีการวัดที่ถูกต้อง เพื่อประมาณค่าคะแนนพัฒนาการ และเพื่อศึกษารูปแบบของพัฒนาการ (Burr and Nesselrode, 1990: 3-4) การวัดคะแนนพัฒนาการในสาขาวิชาการศึกษาเป็นที่สนใจของนักวัดผลมาตั้งแต่ยุคเริ่มต้นของการวัดตั้งแต่ปีค.ศ. 1924 (Embretson, 1991: 415) โดยมุ่งศึกษาภาพพัฒนาการของการเรียนรู้ของผู้เรียน และศึกษาประสิทธิภาพหรือคุณภาพของตัวแปรจัดกระทำที่ให้แก่ผู้เรียนเพื่อใช้ในการประเมินระบบการจัดการศึกษา ว่าบรรลุวัตถุประสงค์หรือไม่อย่างไร (Pike, 1991: 501; Willett, 1994: 671)

วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมมี 2 กลุ่ม กลุ่มแรกเป็นวิธีการวัดคะแนนพัฒนาการจากการวัด 2 ครั้ง เช่นวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน เป็นต้น กลุ่มที่สองเป็นวิธีการวัดคะแนนพัฒนาการจากการวัดตั้งแต่ 2 ครั้งขึ้นไป ซึ่งมีหลายวิธี แต่วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลนี้เป็นพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model: LGCM) ได้รับการยอมรับว่าโมเดลนี้เป็นโมเดลที่ดีที่สุดในการวัดคะแนนพัฒนาการกรณีที่มีการวัดตั้งแต่ 3 ครั้งขึ้นไป (Stoolmiller, 1995) เพราะสามารถบอกได้ว่าแบบแผนของพัฒนาการเป็นรูปแบบใด และประมาณค่าคะแนนพัฒนาการรายบุคคลได้ถูกต้อง รวมทั้งมีความเที่ยงสูง (Willett and Sayer, 1994: 377) แต่มีจุดอ่อนในด้านต้องมีการวัดหลายครั้งและต้องใช้ความรู้ทางสถิติขั้นสูง จึงไม่สะดวกที่จะนำไปใช้ในโรงเรียน และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบมีหลายวิธี เช่น วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ซพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง เป็นต้น จะเห็นได้ว่าวิธีการวัดคะแนนพัฒนาการมีหลายกลุ่มทฤษฎีและหลากหลายวิธีการวัด ซึ่งแต่ละวิธียังไม่มีข้อสรุปที่ชัดเจนว่าวิธีไหนดีที่สุดสำหรับสถานการณ์ใด ผู้วิจัยจึงสนใจที่จะพัฒนาวิธีการวัดคะแนนพัฒนาการที่เป็นวิธีที่ง่าย ไม่เสียเวลาในการวัด และเป็นวิธีที่สะดวกสำหรับครูในโรงเรียนนำไปใช้ปฏิบัติได้ โดยที่มีคุณภาพของคะแนนพัฒนาการจากวิธีการวัดนั้นใกล้เคียงกับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงซึ่งผู้วิจัยใช้เป็นวิธีเกณฑ์ ซึ่งในการวิจัยครั้งนี้ผู้วิจัยได้คัดสรรวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมโดยใช้เกณฑ์ว่าเป็นวิธีที่มีค่าความเที่ยงสูงและค่าความคลาดเคลื่อนต่ำจำนวน 5 วิธี ทฤษฎีการตอบสนองข้อสอบ 2 วิธี แต่ผู้วิจัยพบว่าวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบยังมีจุดอ่อนในด้านของอิทธิพลและความยุติธรรมของคะแนนพัฒนาการสำหรับผู้ที่มีความสามารถเริ่มต้นต่ำ ผู้วิจัย

จึงได้พัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบขึ้น 2 วิธีโดยใช้แนวคิดของศิริชัย กาญจนวาสี (2532) และวินิจ เทือกทอง (2537) ซึ่งใช้ในทฤษฎีการทดสอบแบบดั้งเดิม

วัตถุประสงค์ในการวิจัย

เพื่อพัฒนาและวิเคราะห์เปรียบเทียบคุณภาพคะแนนพัฒนาการระหว่างวิธีการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี ตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี และวิธีที่ผู้วิจัยพัฒนา 2 วิธี

กรอบแนวคิดและสมมติฐานการวิจัย

จากแนวคิดการประมาณค่าคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมที่วัดคะแนนพัฒนาการจากการหาผลต่างระหว่างคะแนนสอบก่อนและหลังเรียน และการปรับแก้วิธีการวัดจากผลต่างระหว่างคะแนนสอบก่อนและหลังเรียน ผู้วิจัยได้คัดสรรมาใช้ในการวิจัย 5 วิธีแต่ละวิธีมีสูตรดังนี้

1. วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (observed difference score method) ในที่นี้ใช้สัญลักษณ์ DS โดย $DS = X_2 - X_1$
2. วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (standard growth method) ในที่นี้ใช้สัญลักษณ์ SC โดย $SC = Z_2 - Z_1$
3. วิธีการวัดคะแนนพัฒนาการลอการิทึมของคะแนนดิบ (logarithm of observed score method) ในที่นี้ใช้สัญลักษณ์ NL โดย $NL = \ln X_2 - \ln X_1$
4. วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (relative growth method) ในที่นี้ใช้สัญลักษณ์ SR โดย $SR = \frac{100 (X_2 - X_1)}{F - X_1}$
5. วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ (residual growth and potential ratio method) ในที่นี้ใช้สัญลักษณ์ SRR โดย $\frac{100R}{F - X_1}$

เมื่อ X_1, X_2	คือ	คะแนนสอบก่อนเรียนและหลังเรียน
Z_1, Z_2	คือ	คะแนนมาตรฐานของคะแนนสอบก่อนเรียนและหลังเรียน
$\ln X_1, \ln X_2$	คือ	ลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียนและหลังเรียน
F	คือ	คะแนนเต็มในการวัด
R	คือ	คะแนนพัฒนาการวิธีพิเศษเหลือ

การวัดคะแนนพัฒนาการตามแนวคิดทฤษฎีการตอบสนองข้อสอบมีการนำค่าพารามิเตอร์ของข้อสอบและความสามารถที่แท้จริงของผู้สอบมาอธิบายความน่าจะเป็นในการตอบข้อสอบถูก

ในรูปของฟังก์ชัน monotonically increasing function จากฟังก์ชันดังกล่าวสามารถนำมาประมาณค่าความสามารถที่แท้จริงของผู้สอบได้ (Bock 1997, Hambleton and Swaminathan, 1985) ต่อมาผู้พัฒนาวิธีการวัดคะแนนพัฒนาการอีกหลายวิธี แต่มี 2 วิธีที่สามารถประมาณค่าคะแนนพัฒนาการรายบุคคลได้ ซึ่งทั้ง 2 วิธีนี้ได้รับการพัฒนาดังนี้ ในปี 1991 Embretson (1991) ได้พัฒนาโมเดล MRMLC สำหรับการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ซพหุมิติ และ ในปี 1998 May and Nicewander (1998) ได้ประมาณค่าคะแนนพัฒนาการด้วยการหาผลต่างระหว่างความสามารถที่แท้จริง โดยประมาณค่าความสามารถที่แท้จริงด้วย EAP (expected a posteriori) เพื่อแก้ปัญหาความคลาดเคลื่อนที่เกิดจากปฏิสัมพันธ์ระหว่างค่าความยากของข้อสอบกับความสามารถเริ่มต้นที่แท้จริงของผู้สอบ ผู้วิจัยจึงใช้ 2 วิธีดังกล่าวมาใช้ในการวิจัยครั้งนี้ นอกจากนี้ผู้วิจัยได้พัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบเพิ่มขึ้นอีก 2 วิธีโดยใช้แนวคิดของศิริชัย กาญจนวาสี (2532) และวินิจ เทือกทอง (2537) รวมเป็น 4 วิธีโดยมีสูตรดังนี้

1. วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง (method of difference between true ability scores) ในที่นี้ใช้สัญลักษณ์ $L\theta$ โดย $L\theta = \theta_2 - \theta_1$

2. วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ซพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (multidimensional Rasch model for learning and change method: MRMLC)

3. วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพันธ์เมื่อเทียบกับศักยภาพการพัฒนา (relative true ability of growth and potential ratio method) พัฒนาโดยใช้แนวคิดของศิริชัย กาญจนวาสี (2532) ในที่นี้ใช้สัญลักษณ์ $S\theta$ โดย $S\theta = \frac{\theta_2 - \theta_1}{Max - \theta_1}$

4. วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพันธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน (relative true ability of growth and initial ability method) พัฒนาโดยใช้แนวคิดของวินิจ เทือกทอง (2537) ในที่นี้ใช้สัญลักษณ์ $V\theta$ โดย $V\theta = \frac{\theta_2 - \theta_1}{Max + \theta_1}$

เมื่อ θ_1, θ_2 คือ ความสามารถที่แท้จริงก่อนเรียนและหลังเรียน

Max คือ ความสามารถสูงสุด

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยกำหนดสมมติฐานวิจัย ดังต่อไปนี้

1. คะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีที่ผู้วิจัยพัฒนาตามทฤษฎีการตอบสนองข้อสอบ มีคุณภาพดีกว่ากลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

2. ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมวิธีการวัด DS และ SR มีคุณภาพดีกว่าวิธีอื่น

3. ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบวิธีการวัด L0 มีคุณภาพดีกว่า MRMLC
4. ภายในกลุ่มวิธีที่ผู้วิจัยพัฒนาวิธีการวัด S0 และ V0 มีคุณภาพไม่ต่างกัน

วิธีดำเนินการวิจัย

ผู้วิจัยได้พัฒนาแบบทดสอบแบบเลือกตอบ 4 ตัวเลือกเพื่อใช้วัดความรู้บางส่วนจำนวน 66 ข้อ ในวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ภาคต้น จำนวน 3 ฉบับที่คู่ขนานกันเพื่อใช้ในการวัดซ้ำ 3 ครั้ง แต่ละฉบับมีค่าความเที่ยงเท่ากับ 0.913, 0.918 และ 0.928 ตามลำดับ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่มๆละ 698 และ 637 คนในภาคเรียนที่ 1 ปีการศึกษา 2543 จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร 12 โรงเรียน

ข้อมูลที่ใช้ในการวิจัยครั้งนี้มี 3 ชุด ชุดแรกเป็นข้อมูลปฐมภูมิที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ที่มีการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างในการวิจัยครั้งนี้ โดยการวัด 3 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ ข้อมูลชุดแรกเป็นการวัดซ้ำด้วยแบบทดสอบฉบับเดิม ข้อมูลชุดที่สองเป็นการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน รวบรวมข้อมูลช่วงระหว่างเดือนมิถุนายน 2543 ถึงเดือนสิงหาคม 2543 ส่วนข้อมูลชุดที่สามเป็นข้อมูลทุติยภูมิซึ่งเป็นฐานข้อมูลจากงานวิจัยของสมถวิล วิจิตรวรรณ (2543) ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ที่วัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง ห่างกันครั้งละ 3 สัปดาห์ โดยเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในภาคเรียนที่ 2 ปีการศึกษา 2542 จำนวน 433 คน ในโรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม ข้อมูลชุดที่ 3 มีการแยกวิเคราะห์เป็นวัดซ้ำ 3 ครั้ง และแบบวัดซ้ำ 5 ครั้ง ข้อมูลทั้ง 3 ชุดมีการตรวจให้คะแนนแบบทวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) ยกเว้นข้อมูลชุดที่สามแบบวัดซ้ำ 5 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิภาคเท่านั้น ดังนั้นในการวิจัยครั้งนี้จึงมีฐานข้อมูลทั้งหมด 7 ฐาน

การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมโดยใช้โปรแกรม SPSS-WIN 10.07 ตามทฤษฎีการตอบสนองข้อสอบโดยใช้โปรแกรม BILOG 3 เมื่อข้อมูลมีการตรวจให้คะแนนแบบทวิภาค และโปรแกรม PARSCALE 2 เมื่อข้อมูลมีการตรวจให้คะแนนความรู้บางส่วน ส่วนคะแนนพัฒนาการที่ประมาณค่าจากโมเดลราล์พหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงใช้โปรแกรม LPCM-WIN 1.0 และคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงใช้โปรแกรม LISREL 8.3 สำหรับการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการด้านความตรงตาม

เกณฑ์สัมพัทธ์ด้วยค่าสถิติความสัมพันธ์แบบ Pearson และ Spearman และด้านความคลาดเคลื่อนของคะแนนพัฒนาการใช้โปรแกรม SPSS-WIN 10.07 ส่วนการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการด้านความตรงตามเกณฑ์สัมพัทธ์ระหว่างกลุ่มวิธีใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นระดับลดหลั่นโดยใช้โปรแกรม HLM-WIN 5.04 ส่วนภายในกลุ่มวิธีใช้สถิติทดสอบ Hotelling โดยใช้โปรแกรม EXCEL

ผลการวิจัย

เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี ในข้อมูลทั้ง 7 ฐาน พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีคุณภาพไม่แตกต่างกันยกเว้นข้อมูลชุดที่ 3 ที่มีการตรวจให้คะแนนความรู้บางส่วน ดังรายละเอียดผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพัทธ์ด้วย HLM และค่าความคลาดเคลื่อนในตารางที่ 1 และ 2 ตามลำดับ

ส่วนผลการเปรียบเทียบภายในกลุ่มวิธีการวัดด้วยสถิติทดสอบ t ของ Hotelling (1940) พบว่าคะแนนพัฒนาการจากวิธีการวัดส่วนใหญ่มีค่าความตรงตามเกณฑ์สัมพัทธ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยพบว่าภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม วิธีการวัด DS และ SR มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และวิธีการวัด SRR มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3 ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าวิธีการวัด MRMLC มีค่าคุณภาพสูงกว่าวิธีการวัด L0 และผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าวิธีการวัด V0 มีคุณภาพสูงกว่า S0 เมื่อมีการตรวจให้คะแนนแบบทวิภาค ดังรายละเอียดค่าความตรงตามเกณฑ์สัมพัทธ์ และค่าความคลาดเคลื่อนในตารางที่ 1 และ 2 ตามลำดับ

◆ การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบ ◆
แบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 1 ค่าความตรงตามเกณฑ์สัมพัทธ์ของข้อมูลแต่ละชุดและการตรวจสอบความแตกต่างของค่า ความตรงตามเกณฑ์สัมพัทธ์ระหว่างกลุ่มวิธีด้วย HLM

ฐานข้อมูลชุดที่	repeat		counterbalance		equivalence 1		equivalence 2
	dichotomous	partial credit	dichotomous	partial credit	dichotomous	partial credit	dichotomous
	1	2	3	4	5	6	7
	ค่าความตรงตามเกณฑ์สัมพัทธ์						
วิธีCTT							
DS	0.9658**	0.9929**	0.9492**	0.9535**	0.7227**	0.7155**	0.7061**
SC	0.8000**	0.8242**	0.8579**	0.8960**	0.6602**	0.7085**	0.5879**
NL	0.7999**	0.8755**	0.8082**	0.8604**	0.6136**	0.6043**	0.5630**
SR	0.9583**	0.9459**	0.8738**	0.8071**	0.7108**	0.7528**	0.7228**
SRR	0.5478**	0.2329**	0.4304**	0.2864**	0.9733**	0.9532**	0.7642**
วิธีIRT							
L θ	0.7797**	0.8067**	0.6967**	0.7439**	0.5593**	0.4358**	0.5420**
MRMLC	0.9480**	-	0.9418**	-	0.6761**	-	0.6610**
วิธีที่พัฒนา							
s θ	0.6757**	0.7925**	0.6201**	0.7314**	0.4411**	0.4244**	0.4599**
v θ	0.7879**	0.8072**	0.6959**	0.6638**	0.5850**	0.4065**	0.5536**
	ผลการวิเคราะห์ด้วย HLM						
SD ของค่าความตรง	0.002	0.002	0.010	0.009	0.090	0.179	0.067
χ^2	0.940	0.027	1.063	0.132	5.222	19.465	5.840
p	>0.500	>0.500	>0.500	>0.500	0.072	0.000**	0.052

** P < 0.01

ตารางที่ 2 ค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนของข้อมูลแต่ละชุด

ฐานข้อมูลชุดที่	repeat		counterbalance		equivalence 1		equivalence 2
	dichotomous	partial credit	dichotomous	partial credit	dichotomous	partial credit	dichotomous
	1	2	3	4	5	6	7
วิธี CTT							
DS	0.2612	0.1194	0.3185	0.3048	0.7455	0.7558	0.7674
SC	0.6320	0.5924	0.5326	0.4557	0.8252	0.7651	0.9087
NL	0.6323	0.4986	0.6189	0.5280	0.8811	0.8920	0.9337
SR	0.2886	0.3288	0.5020	0.6205	0.7615	0.7045	0.7464
SRR	0.9503	1.2377	1.0664	1.1937	0.2317	0.3070	0.9993
วิธี IRT							
L θ	0.6633	0.6213	0.7782	0.7151	0.9375	1.0625	0.9627
MRMLC	0.3235		0.3449		0.8063		0.8237
วิธีที่พัฒนา							
S θ	0.8048	0.6438	0.8709	0.7324	1.0581	1.0731	1.0481
V θ	0.6508	0.6205	0.7792	0.8193	0.9078	1.0902	0.9486

การอภิปรายผลการวิจัย

1 ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างกลุ่มวิธี ผลการวิจัยครั้งนี้ขัดแย้งกับสมมติฐานวิจัยข้อ 1 กล่าวคือ ผลการวิจัยพบว่าคุณภาพของวิธีการวัดคะแนนพัฒนาการทุกกลุ่มวิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ยกเว้นข้อมูลชุดที่ 3 ที่มีการตรวจให้คะแนนความรู้บางส่วนที่พบว่ากลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ซึ่งอธิบายได้ด้วยเหตุผล 2 ประการ ประการแรก กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมมีการประมาณค่าคะแนนพัฒนาการจากคะแนนดิบก่อนเรียนและหลังเรียนโดยตรง เช่นเดียวกับการประมาณค่าคะแนนพัฒนาการตามวิธีที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ (Meredith and Tisak, 1990; McArdle and Hamagami, 1995) ซึ่งมีผลทำให้ค่าความตรงตามเกณฑ์สัมพัทธ์มีค่าสูง ในขณะที่คะแนนพัฒนาการที่ได้จากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบเป็นการประมาณค่าจากฟังก์ชันของคะแนนความสามารถที่แท้จริงของผู้สอบซึ่งขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ (Embretson, 1991; May and Nicewander, 1998) ซึ่งอาจมีผลทำให้คะแนนพัฒนาการที่ได้มีความสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์ในระดับต่ำได้

ประการที่สองคะแนนพัฒนาการที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ได้มาจากการประมาณค่าด้วยโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง คะแนนพัฒนาการที่ได้ตามวิธีนี้อาจจะไม่สอดคล้องและอาจจะเป็นคะแนนพัฒนาการคนละแบบกับคะแนนพัฒนาการที่เป็นความสามารถที่แท้จริงตามทฤษฎีการตอบสนองข้อสอบก็ได้

2 ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างวิธีการวัดภายในกลุ่มวิธี ผลการวิจัยกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมสอดคล้องกับสมมติฐานวิจัยข้อ 2 เพียงบางส่วน กล่าวคือ ผลการวิจัยพบว่าวิธีการวัด DS และ SR มีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่นเฉพาะข้อมูล 2 ชุดแรก และสำหรับข้อมูลชุดที่ 3 วิธีการวัด SRR มีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่น ทั้งนี้อาจเป็นเพราะว่าข้อมูล 2 ชุดแรกมีค่าสุดโต่งมากกว่าข้อมูลชุดที่ 3 และค่าสุดโต่งส่งผลต่อการประมาณค่าคะแนนพัฒนาการด้วยสมการทำนายทำให้กราฟสมการการทำนายเบี่ยงเบนไปจากกราฟสมการทำนายเมื่อไม่มีอิทธิพลจากค่าสุดโต่ง (Hair, Anderson, and Black, 1998: 185) ผลการวิจัยกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบขัดแย้งกับสมมติฐานวิจัยข้อ 3 กล่าวคือผลการวิจัยพบว่าวิธีการวัด L θ มีคุณภาพต่ำกว่าวิธีการวัด MRMLC ผลการวิจัยที่ได้ดังกล่าวนี้อาจอธิบายได้ว่าเนื่องจากวิธีการประมาณค่าพารามิเตอร์ของทั้ง 2 วิธีแตกต่างกันวิธีการวัด L θ ประมาณค่าพารามิเตอร์ด้วยวิธี EAP (expected a posteriori estimator) ส่วนวิธีการวัด MRMLC ประมาณค่าพารามิเตอร์ด้วยวิธี ML (maximum likelihood estimator) การประมาณค่าพารามิเตอร์ด้วยวิธี ML มีจุดอ่อนในด้านที่ไม่สามารถประมาณค่าพารามิเตอร์ของผู้ที่ได้คะแนนเป็นศูนย์และผู้ที่ได้คะแนนเต็ม (Mislevy and Bock, 1990: 1-13) ดังนั้นคะแนนพัฒนาการที่ประมาณค่าด้วยวิธีการวัด MRMLC จึงต้องมีการตัดคะแนนที่เป็นศูนย์และคะแนนเต็มออกไป จึงทำให้กลุ่มตัวอย่างที่เหลืออยู่ไม่มีอิทธิพลพื้นและอิทธิพลเพดาน คะแนนพัฒนาการจึงมีค่าใกล้เคียงกับคะแนนพัฒนาที่เป็นเกณฑ์ ทำให้คุณภาพของวิธีการวัด MRMLC มีค่าสูงกว่าวิธีการวัด L θ ซึ่งมีอิทธิพลพื้นและอิทธิพลเพดานรวมอยู่ด้วย สำหรับผลการวิจัยกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาสอดคล้องกับสมมติฐานวิจัยข้อ 4 เพียงบางส่วน กล่าวคือผลการวิจัยพบว่าวิธีการวัด V θ มีคุณภาพสูงกว่าวิธีการวัด S θ ในกรณีที่ข้อมูลมีการตรวจให้คะแนนแบบทวิภาค และวิธีการวัด V θ กับ S θ มีคุณภาพแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ เมื่อข้อมูลมีการตรวจให้คะแนนความรู้บางส่วน ผลการวิจัยที่ได้ดังกล่าวนี้อาจอธิบายได้ว่าเกิดจากวิธีการตรวจให้คะแนนต่างกันมีผลทำให้ลักษณะการแจกแจงของข้อมูลแตกต่างกัน การวิจัยครั้งนี้ พบว่าลักษณะการแจกแจงของข้อมูลทุกชุดจากการสอบทุกครั้งมีลักษณะค่อนข้างที่จะเบ้ขวา และลักษณะความเบ้ขวาลดลงในการวัดครั้งหลังแต่ปริมาณลักษณะการลดลงของความเบ้ในข้อมูลชุดที่มีการตรวจให้คะแนนแบบทวิภาคมีน้อยกว่าในข้อมูลชุดที่มีการ

ตรวจให้คะแนนความรู้อย่างส่วน อย่างไรก็ตามข้อมูลในการวิจัยครั้งนี้ยังไม่อาจสรุปได้อย่างแน่ชัดว่าวิธีการวัด V0 มีคุณภาพไม่ต่างจากวิธีการวัด S0 เนื่องจากวิธีการตรวจให้คะแนน ควรที่จะต้องมีการวิจัยตรวจสอบโดยใช้การวิจัยแบบจำลอง (simulation) ข้อมูลที่มีปริมาณการลดลงของความเบ้ในการวัดครั้งหลัง ๆ ด้วย

ข้อเสนอแนะ

ผลการวิจัยครั้งนี้พบว่า ผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ในโรงเรียนโครงการขยายโอกาสการศึกษาขั้นพื้นฐาน และระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนในสังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร มีพัฒนาการค่อนข้างต่ำ ส่วนใหญ่จะมีพัฒนาการทางคณิตศาสตร์ต่ำกว่าค่าเฉลี่ย ดังนั้นหน่วยงานที่เกี่ยวข้องน่าจะได้นำทางแก้ไข เพราะวิชาคณิตศาสตร์เป็นวิชาที่สำคัญในการที่จะพัฒนาเยาวชนของประเทศไทยให้มีความสามารถและมีศักยภาพที่จะแข่งขันกับประเทศอื่นได้

ในระดับประถมศึกษาและมัธยมศึกษา โรงเรียนควรที่จะใช้วิธีการวัดคะแนนพัฒนาการของนักเรียนโดยจัดให้มีการวัด 2 ครั้ง ก่อนและหลังเรียน ด้วยเครื่องมือชุดเดิม ทั้งนี้อาจจัดสอบแบบจตุรัสละตินด้วยเครื่องมือการวัดที่คู่ขนานกันเพื่อแก้ปัญหาในด้านการจำข้อสอบได้ในระดับมัธยมศึกษา (Embretson, 1991) และควรใช้วิธีการวัดคะแนนพัฒนาการด้วยวิธีการวัด DS และ SR และเนื่องจากในปัจจุบันครูส่วนใหญ่มีศักยภาพในด้านการใช้คอมพิวเตอร์ หากเป็นไปได้ควรใช้วิธีการวัด MRMLC

การวิจัยครั้งนี้มีการกำหนดขอบเขตการวิจัย โดยพิจารณาคุณภาพของวิธีการวัดคะแนนพัฒนาการด้วยความตรงตามเกณฑ์สัมพันธและความคลาดเคลื่อนเท่านั้น นอกจากนี้ยังกำหนดขอบเขตการวิเคราะห์เฉพาะข้อมูลเพียง 7 รฐานซึ่งมีผลทำให้การสรุปผลการวิจัยยังไม่หนักแน่น หากมีการขยายขอบเขตการวิจัยให้ครอบคลุมเกณฑ์ในการตรวจสอบคุณภาพและการจำลองข้อมูล (simulation) ให้ครอบคลุมลักษณะข้อมูลทุกรูปแบบ จะทำให้ได้ผลการวิจัยที่ชัดเจนยิ่งขึ้น

เอกสารอ้างอิง

ภาษาไทย

- นงลักษณ์ วิรัชชัย. (2542). **โมเดลลิสม์เรล: สถิติวิเคราะห์สำหรับการวิจัย**. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- วินิจ เทือกทอง. (2537). การศึกษาเปรียบเทียบประสิทธิผลของการคำนวณคะแนนเพิ่มวิธีต่าง ๆ ด้วยระเบียบวิธีกรมอนติคาร์โล. **วิทยานิพนธ์ปริญญาโทมหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร**.
- ศิริชัย กาญจนวาสี. (2544). **ทฤษฎีการทดสอบแบบดั้งเดิม (CLASSICAL TEST THEORY)**. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สมถวิล วิจิตรวรรณ. (2543). **การเปรียบเทียบผลที่ได้จากการวัดพัฒนาการตัวแปรเอกนาม และพหุนามจากการใช้ 3 โมเดล คือ โมเดลพัฒนาการที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดล กึ่งซิมเพลกซ์**. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- อิทธิพงษ์ ตั้งสกุลเรืองไฉ. (2541). **การเปรียบเทียบประสิทธิผลของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 4 รูปแบบในการศึกษาการเปลี่ยนแปลงระยะยาวของการพัฒนาทางกาย และผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนประถมศึกษา**. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- อรุณี อ่อนสวัสดิ์. (2537). **การพัฒนาวิธีการวัดการเปลี่ยนแปลงการเรียนรู้**. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- Bryk, A. S. and Raudenbush, S. W. (1992). **Hierarchical linear models**. California: Sage Publications, Inc.
- Bryk, A. S., Raudenbush, S. W. and Congdon, R. T. (1994). **HLM: Hierarchical linear modeling with the HLM/2L and HLM/3L programs**. Chicago: Scientific Software, Inc.
- Embretson, S. E. (1991b). A multidimensional latent trait model for measuring learning and change. **Psychometrika**. 56: 495-515.
- Fischer, G. H. and Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. **Psychometrika**. 56: 637-651.
- Fischer, G. H. and Ponocny, I. (1994). An extension of the partial credit model with an application to the measurement of change. **Psychometrika**. 59: 177-192.

- Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998). **Multivariate data analysis.** (5th ed). NJ: Prentice-Hall, Inc.
- May, K. and Nicewander, W. A. (1998). Measuring change conventionally and adaptively. **Educational and Psychological Measurement.** 58: 882-897.
- Mislevy, R. J. and Bock, R. D. (1990). **BILOG 3 : Item analysis and test scoring with binary logistic models.** Chicago: Scientific Software, Inc.
- Muraki, E. and Bock, R. D. (1993). **PARSCALE 2: IRT Based test scoring and item analysis for graded open-ended exercises and performance tasks.** Chicago: Scientific Software, Inc.
- Ponocny, I. And Ponocny, E. S. (1997). Applications of the program LPCM in the field of measuring change. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective Measurement IV: Theory into Practice,** pp. 209-222. Greenwich, CT: Ablex.
- Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: growth and change from childhood to early adolescence. **Multivariate Behavioral Research.** 29: 263-288.
- Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In John M. Gottman (Ed.), **The Analysis of Change,** pp. 103-138. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.