

การเปรียบเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน: การเปรียบเทียบ คุณภาพระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบ

สุนนา ศรีอบเชย
เสรี ชัดเข้ม และ สมโภชน์ อเนกสุข
มหาวิทยาลัยบูรพา

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของการเปรียบเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน โดยการเปรียบเทียบความคลาดเคลื่อนและความเพียงพอของการเปรียบเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 4, 5 และ 6 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 3 ปีการศึกษา 2548 จำนวน 1,260 คน จำแนกเป็น 2 กลุ่ม ได้แก่ กลุ่มเปรียบเทียบคะแนนจำนวน 900 คน และกลุ่มสอบทานผลจำนวน 360 คน เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ จำนวน 4 ฉบับ แบ่งเป็นแบบทดสอบสำหรับกลุ่มเปรียบเทียบคะแนนตามระดับชั้น จำนวน 3 ฉบับ ๆ ละ 25 ข้อ ซึ่งมีข้อสอบร่วมระหว่างระดับชั้น ฉบับละ 5 ข้อ และแบบทดสอบสำหรับกลุ่มสอบทานผล 1 ฉบับ จำนวน 65 ข้อ

ผลการวิจัยปรากฏว่า

1. ความคลาดเคลื่อนของวิธีเชิงเส้นตรงมีค่าน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2. คำนวณความแตกต่าง (C) ของวิธีเชิงเส้นตรง มีค่าน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า วิธีเชิงเส้นตรงมีความเพียงพอของการเปรียบเทียบคะแนนอยู่ในระดับที่น่าพอใจมากกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

สรุปได้ว่าการเปรียบเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายในกับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิธีเชิงเส้นตรงมีคุณภาพดีกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

ส่วนหนึ่งของวิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีวิทยาการศึกษา มหาวิทยาลัยบูรพา ปีการศึกษา 2550 อาจารย์ที่ปรึกษาหลัก รศ.ดร. เสรี ชัดเข้ม และอาจารย์ที่ปรึกษาร่วม ผศ.ดร.สมโภชน์ อเนกสุข

Vertical Equating Using Internal Anchor Tests: A Comparative of the Relative Quality of Linear Equating and IRT Equating Methods

Sumana Sriobchey

Seree Chadcham and Sompoch Anegasukha

Barapha University, Thailand

ABSTRACT

The purpose of this research was to study the quality of vertical equating using internal anchor items by comparing the error and adequacy of equating between linear and IRT two-parameter methods. The sample consisted of 1,260 Grade 4, 5 and 6 students in schools under the Chon Buri Education Service Area Office, Area III of academic year 2005. This sample was divided into two groups; 900 for the equating group, and 360 for the cross-validation group. The research instruments were four mathematics achievement tests, three for the equating group with 25 items (with five internal anchor items), and one for the cross-validation group with 65 items.

The results were as follow:

- 1. The error of the linear equating method was found to be less than that of the IRT equating method at the .01 level of statistical significance.*
- 2. The discrepancy index (C) of the linear equating method was found to be less than the IRT equating method at the .01 level of statistical significance, indicating that the linear equating method yielded results that could be more acceptable than those of the IRT two-parameter equating method.*

With regard to vertical equating using internal anchor items with achievement tests, the linear equating method was found to be superior to the IRT two-parameter equating method.

Based on a master's thesis completed at the Department of Educational Research and Measurement, Burapha University, under the supervision of Assoc. Prof. Seree Chadcham, Ph.D., and Assist. Prof. Sompoch Anegasukha, Ed.D.

ความนำ

การปรับเทียบคะแนนระหว่างแบบทดสอบ (Test equating) เป็นการปรับคะแนนจากแบบทดสอบฉบับหนึ่งให้อยู่บนมาตราเดียวกับแบบทดสอบอีกฉบับหนึ่งโดยใช้วิธีการทางสถิติ จำแนกเป็น 2 ลักษณะ ได้แก่ 1) การปรับเทียบคะแนนตามแนวนอนหรือแนวราบ (Horizontal equating) เป็นการปรับเทียบคะแนนในกรณีที่แบบทดสอบที่นำมาปรับเทียบกันมีระดับความยากใกล้เคียงกันและกลุ่มผู้สอบมีความสามารถใกล้เคียงกันหรืออยู่ในระดับชั้นเดียวกัน 2) การปรับเทียบคะแนนตามแนวตั้งหรือแนวตั้ง (Vertical equating) เป็นการปรับเทียบคะแนนในกรณีที่แบบทดสอบที่นำมาปรับเทียบกันมีระดับความยากต่างกันและกลุ่มผู้สอบมีความสามารถต่างกันหรืออยู่ต่างระดับชั้นกัน (Hambleton & Swanminathan, 1985)

วิธีการปรับเทียบคะแนนระหว่างแบบทดสอบทั้งแนวนอนและแนวตั้งที่ใช้กันแพร่หลาย ได้แก่ วิธีอีควิเปอร์เซ็นต์ไทล์ (Equipercentile methods) วิธีเชิงเส้นตรง (Linear methods) วิธีทฤษฎีการตอบสนองข้อสอบ (Item response theory methods) ซึ่งแต่ละวิธีก็มีเงื่อนไขในการปรับเทียบคะแนนที่แตกต่างกัน เช่น ลักษณะของกลุ่มผู้สอบ ความแตกต่างของระดับความสามารถของกลุ่มผู้สอบ ลักษณะการกระจายของคะแนน เป็นต้น จากการศึกษาของแองกอฟฟ์ (Angoff, 1984: 87) พบว่า ถ้ารูปทรงการแจกแจงคะแนนของแบบทดสอบต่างชุดแตกต่างกันแล้ว การปรับเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์จะให้ผลดีกว่าการปรับเทียบคะแนนเชิงเส้นตรง แต่หากรูปทรงการแจกแจงคะแนนของแบบทดสอบคล้ายคลึงกัน การปรับเทียบคะแนนเชิงเส้นตรงจะให้ผลดีกว่าการปรับเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์ เพราะมีความเป็นปรนัยมากกว่า ส่วนลอร์ด (Lord, 1980: 199) ได้ศึกษาการปรับเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบพบว่า มีคุณสมบัติของความไม่แปรผันไปตามกลุ่มตัวอย่างที่นำมาใช้ปรับเทียบมาตรา จึงให้ผลดีที่สุด แม้ว่าในกรณีที่การแจกแจงของคะแนนจากแบบทดสอบต่างชุดแตกต่างกัน ซึ่งสอดคล้องกับงานวิจัยของโคลน (Kolen, 1981) ที่พบว่า การปรับเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบมีความเหมาะสมที่สุด รองลงมาคือวิธีการปรับเทียบคะแนนเชิงเส้นตรงสำหรับแบบทดสอบที่มีความยากต่างกัน

การปรับเทียบคะแนนเชิงเส้นตรงมีข้อดีดังนี้

1. การปรับเทียบคะแนนเชิงเส้นตรงเป็นวิธีการแปลงคะแนนที่ง่าย และสะดวกในการใช้มากที่สุด (Kolen & Whitney, 1982; Petersen, Cook & Stocking, 1983; Hill & Subhiyah, 1988)
2. การปรับเทียบคะแนนเชิงเส้นตรงให้ผลดีกว่าวิธีอื่นกับการวัดผลสัมฤทธิ์ทางการเรียน (Petersen, Cook & Stocking, 1983) เนื่องจากการวัดผลสัมฤทธิ์ทางการเรียนในชั้นเรียนมีขอบเขตเนื้อหาที่จำกัดตามหลักสูตร แบบทดสอบชุดใหม่ที่สร้างขึ้นจะประกอบด้วยข้อสอบที่มีลักษณะต่าง ๆ ที่ใกล้เคียงกับแบบทดสอบที่สร้างขึ้นชุดก่อนๆ เมื่อแบบทดสอบวัดเนื้อหาเดียวกัน วิธีการปรับเทียบคะแนนเชิงเส้นตรงจึงให้ผลดีกว่าวิธีอื่น (Kolen & Whitney, 1982)

3. การเปรียบเทียบคะแนนเชิงเส้นตรงยังคงใช้ได้ดีกับกลุ่มตัวอย่างที่มีขนาดเล็ก (Kolen & Whitney, 1982) ซึ่งผู้สอนสามารถนำไปใช้ปฏิบัติได้จริงในสภาพการเรียนการสอน โดยเฉพาะในชั้นเรียนที่มีจำนวนนักเรียนไม่มากนัก

ส่วนการเปรียบเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบ มีข้อดีดังนี้

1. ผลการเปรียบเทียบคะแนนไม่ขึ้นอยู่กับระดับความสามารถของกลุ่มที่นำไปเปรียบเทียบ
2. ผลการเปรียบเทียบคะแนนไม่ได้รับผลกระทบจากแบบทดสอบที่มีความยากต่างกัน

นอกจากนี้การเปรียบเทียบคะแนนยังขึ้นอยู่กับข้อมูลที่น่าสนใจ ซึ่งแองกอฟฟ์ (1984: 105) ได้จำแนกการเปรียบเทียบมาตราเป็น 2 กลุ่มใหญ่ ๆ ได้แก่ การเปรียบเทียบมาตราโดยใช้แบบทดสอบร่วม และการเปรียบเทียบมาตราโดยไม่ใช้แบบทดสอบร่วม แบบทดสอบร่วมที่ใช้ในการเปรียบเทียบคะแนนเป็นกลุ่มข้อสอบที่กำหนดให้ผู้สอบสองกลุ่มซึ่งทำแบบทดสอบคนละชุดต้องทำเหมือนกัน แบบทดสอบร่วมสามารถแยกหรือรวมอยู่ในแบบทดสอบที่ต้องการเปรียบเทียบคะแนนก็ได้ (Angoff, 1984: 107) แบบทดสอบร่วมที่แยกออกจากแบบทดสอบที่ต้องการเปรียบเทียบคะแนน โดยไม่รวมคะแนนส่วนนี้เข้าไปในผลการสอบรายบุคคล เรียกว่า แบบทดสอบร่วมภายนอก (External anchor test) ส่วนแบบทดสอบร่วมที่รวมคะแนนส่วนนี้เข้าไปในแบบทดสอบที่ต้องการเปรียบเทียบคะแนน เรียกว่า แบบทดสอบร่วมภายใน (Internal anchor test) การเปรียบเทียบคะแนนโดยไม่ใช้แบบทดสอบร่วมเป็นรูปแบบที่มีเงื่อนไขหลายประการ และยังมีข้อได้เปรียบในทางปฏิบัติ จึงนิยมใช้การเปรียบเทียบคะแนนโดยใช้แบบทดสอบร่วม

จากการศึกษาเอกสารและงานวิจัยที่ผ่านมา (Johanson, 1988; Harris, 1991; Ayerve, 1992; Leung, 2003, พรพิมล นาคเวช, 2537; เกตรินทร์ หมัดอะดัม, 2545) ยังไม่อาจสรุปได้ชัดเจนว่าวิธีการเปรียบเทียบคะแนนตามแนวทางวิธีใดเหมาะสมที่จะนำไปใช้วัดผลสัมฤทธิ์ทางการเรียนมากที่สุด เนื่องจากการเปรียบเทียบคะแนนวิธีหนึ่งอาจให้ผลดีในสถานการณ์หนึ่งและอาจให้ผลด้อยในอีกสถานการณ์หนึ่ง ดังนั้นจึงควรเลือกใช้วิธีการเปรียบเทียบคะแนนที่เหมาะสมเพื่อให้ได้ผลการเปรียบเทียบคะแนนที่ถูกต้องและเชื่อถือได้ ผู้วิจัยจึงสนใจที่จะศึกษาคุณภาพของการเปรียบเทียบคะแนนตามแนวทางโดยใช้แบบทดสอบร่วมภายใน โดยเลือกเปรียบเทียบวิธีการเปรียบเทียบคะแนนตามแนวทาง 2 วิธี คือวิธีเชิงเส้นตรงและวิธีทฤษฎีการตอบสนองข้อสอบ แล้วพิจารณาว่าวิธีใดมีความคลาดเคลื่อนและความเพียงพองของการเปรียบเทียบคะแนนน้อยกว่ากัน เพื่อจะได้ทราบวิธีการเปรียบเทียบคะแนนที่มีคุณภาพและเหมาะสมกับสถานการณ์ของการเปรียบเทียบคะแนนตามแนวทาง และสามารถนำไปใช้เป็นแนวทางในการตัดสินใจเลือกวิธีการเปรียบเทียบคะแนนของนักเรียนที่ใช้แบบทดสอบต่างฉบับวัดเนื้อหาเดียวกัน แต่ต่างระดับชั้นกันได้อย่างเหมาะสม และมีความถูกต้องมากยิ่งขึ้น อีกทั้งยังทำให้เกิดความยุติธรรมแก่ผู้สอบทุกคนในการเปรียบเทียบและตีความหมายคะแนนระหว่างแบบทดสอบ

วัตถุประสงค์การวิจัย

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายในระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ในเรื่องต่อไปนี้

1. ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน โดยวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์
2. ค่าความเพียงพอของการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายในโดยวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

สมมติฐานการวิจัย

จากงานวิจัยของเหลียง (Leung, 2003) พบว่า การใช้วิธีการปรับเทียบคะแนนโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์และวิธีเชิงเส้นตรง โดยใช้แบบทดสอบร่วมภายในกับการปรับเทียบคะแนนตามแนวตั้งนั้นได้ผลดีเช่นเดียวกัน เนื่องจากมีค่าสัมประสิทธิ์สหสัมพันธ์ที่สูงและเข้าใกล้ 1.00 และจากงานวิจัยของโคเลน (Kolen, 1981) พบว่า แบบทดสอบที่มีความยากต่างกัน และผู้สอบมีความสามารถต่างกัน หรืออยู่ต่างระดับชั้นกันแล้ว วิธีการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบมีความเหมาะสมที่สุดรองลงมาคือวิธีการปรับเทียบเชิงเส้นตรง จึงตั้งสมมติฐานการวิจัยดังนี้

1. ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน ด้วยวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้นตรง
2. ค่าความเพียงพอของการปรับเทียบคะแนนตามแนวตั้ง โดยใช้แบบทดสอบร่วมภายใน ด้วยวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้นตรง

นิยามศัพท์เฉพาะ

ค่าความคลาดเคลื่อนของการปรับเทียบคะแนน (Error of equating) หมายถึง ความคลาดเคลื่อนของค่าเฉลี่ยที่แสดงความแตกต่างที่เกิดขึ้นระหว่างคะแนนที่ได้จากตารางปรับเทียบคะแนนกับคะแนนที่ได้จากการสอบของกลุ่มตัวอย่างสอบทานผล

ค่าความเพียงพอของการปรับเทียบคะแนน (Adequacy of equating) หมายถึง ค่าที่แสดงความแตกต่างของการปรับเทียบคะแนนจากแบบทดสอบฉบับหนึ่งไปสู่แบบทดสอบอีกฉบับหนึ่งตามวิธีการปรับเทียบคะแนน การตัดสินความเพียงพอของการปรับเทียบคะแนนใช้การเปรียบเทียบดัชนีความแตกต่าง (Discrepancy index: C) ตามหลักการประเมินของปีเตอร์เซนและคณะ (Petersen et al., 1982)

วิธีการวิจัย

แบบแผนการวิจัย

การเปรียบเทียบคะแนนครั้งนี้เก็บรวมข้อมูลโดยใช้แบบทดสอบร่วมภายในกับกลุ่มผู้สอบที่มีความสามารถต่างกัน 3 กลุ่ม ได้แก่ นักเรียนระดับชั้นประถมศึกษาปีที่ 4, 5 และ 6 โดยกลุ่มตัวอย่างเปรียบเทียบมาตรฐานทำแบบทดสอบในระดับชั้นของตนเองเพียง 1 ฉบับ ซึ่งในแบบทดสอบมีข้อสอบร่วมระหว่างชั้น ๆ ละ 5 ข้อ ส่วนในกลุ่มตัวอย่างสอบทานผล ต้องทำแบบทดสอบทั้ง 3 ฉบับ ซึ่งผู้วิจัยนำมารวมไว้เป็นฉบับเดียวกัน รวมข้อสอบที่ต้องทำทั้งหมด 65 ข้อ ดังตารางที่ 1 และ 2

ตารางที่ 1 แบบทดสอบและแบบทดสอบร่วมระหว่างระดับชั้นของกลุ่มตัวอย่างเปรียบเทียบคะแนน

กลุ่มตัวอย่าง เปรียบเทียบคะแนน	แบบทดสอบ				
	แบบทดสอบ ระดับ ป.4	แบบทดสอบร่วม ระหว่าง ป.4 กับ ป.5	แบบทดสอบ ระดับ ป.5	แบบทดสอบร่วม ระหว่าง ป.5 กับ ป.6	แบบทดสอบ ระดับ ป.6
ป.4	20 ข้อ	5 ข้อ	-	-	-
ป.5	-	5 ข้อ	15 ข้อ	5 ข้อ	-
ป.6	-	-	-	5 ข้อ	20 ข้อ

ตารางที่ 2 แบบทดสอบและแบบทดสอบร่วมระหว่างระดับชั้นของกลุ่มตัวอย่างสอบทานผล

กลุ่มตัวอย่าง สอบทานผล	แบบทดสอบ				
	แบบทดสอบ ระดับ ป.4	แบบทดสอบ ร่วมระหว่าง ป.4 กับ ป.5	แบบทดสอบ ระดับ ป.5	แบบทดสอบ ร่วมระหว่าง ป.5 กับ ป.6	แบบทดสอบ ระดับ ป.6
ป.4	20 ข้อ	5 ข้อ	15 ข้อ	5 ข้อ	20 ข้อ
ป.5	20 ข้อ	5 ข้อ	15 ข้อ	5 ข้อ	20 ข้อ
ป.6	20 ข้อ	5 ข้อ	15 ข้อ	5 ข้อ	20 ข้อ

กลุ่มตัวอย่าง

กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นประถมศึกษาปีที่ 4, 5 และ 6 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 3 แบ่งออกเป็น 2 กลุ่ม คือ กลุ่มตัวอย่างเปรียบเทียบคะแนน และกลุ่มตัวอย่างสอบทานผล

ซึ่งเบรนนันและโคเลน (Brennan & Kolen, 1987) ได้แนะนำว่ากลุ่มตัวอย่างที่ใช้ในการปรับเทียบคะแนนไม่ควรน้อยกว่า 400 คน และโคเลนและวิทนี (Kolen & Whitney, 1982) ได้เสนอว่า กลุ่มตัวอย่างที่ใช้สอบทานผลควรมีขนาดไม่น้อยกว่า 10 เปอร์เซนต์ของกลุ่มตัวอย่างปรับเทียบคะแนน ในงานวิจัยนี้ผู้วิจัยใช้กลุ่มตัวอย่างปรับเทียบคะแนนจำนวน 900 คน และกลุ่มตัวอย่างสอบทานผลจำนวน 360 คน (40 เปอร์เซนต์ของกลุ่มปรับเทียบคะแนน) โดยได้เลือกโรงเรียนที่อยู่ในเครือเดียวกันและใช้แผนการสอนที่คล้ายคลึงกันจำนวน 3 โรงเรียนและนักเรียนที่เป็นกลุ่มตัวอย่างได้มาโดยวิธีการสุ่มอย่างง่ายสุ่มมาโรงเรียนละเท่า ๆ กัน

เครื่องมือการวิจัย

เครื่องมือการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนชนิด 4 ตัวเลือก วิชาคณิตศาสตร์ เรื่อง ทศนิยม สำหรับชั้นประถมศึกษาปีที่ 4, 5 และ 6 เนื่องจากเนื้อหาเรื่องทศนิยมมีการเรียนในทุกระดับชั้น ทำให้นักเรียนในแต่ละระดับชั้นไม่เกิดความได้เปรียบเสียเปรียบกัน ผู้วิจัยได้สร้างแบบทดสอบระดับชั้น ๆ ละ 25 ข้อ แบบทดสอบที่ใช้ปรับเทียบคะแนนแต่ละฉบับมีข้อสอบร่วมภายใน ซึ่งแองกอฟฟ์ (1984) ได้เสนอแนะว่าควรมีไม่ต่ำกว่าร้อยละ 20 ของจำนวนข้อสอบในแบบทดสอบ จึงได้กำหนดให้มีข้อสอบร่วมภายในจำนวน 5 ข้อในแบบทดสอบระดับชั้นที่ติดกัน

1. แบบทดสอบที่ใช้สอบกับกลุ่มตัวอย่างปรับเทียบคะแนน มีดังนี้

1.1 แบบทดสอบสำหรับชั้นประถมศึกษาปีที่ 4 มีข้อสอบเฉพาะ 20 ข้อ และข้อสอบร่วมกับชั้นประถมศึกษาปีที่ 5 จำนวน 5 ข้อ รวมเป็น 25 ข้อ มีค่าความเที่ยงเท่ากับ .87

1.2 แบบทดสอบสำหรับชั้นประถมศึกษาปีที่ 5 มีข้อสอบเฉพาะ 15 ข้อ และข้อสอบร่วมกับชั้นประถมศึกษาปีที่ 4 จำนวน 5 ข้อ และข้อสอบร่วมกับชั้นประถมศึกษาปีที่ 6 จำนวน 5 ข้อ รวมเป็น 25 ข้อ มีค่าความเที่ยงเท่ากับ .86

1.3 แบบทดสอบสำหรับชั้นประถมศึกษาปีที่ 6 มีข้อสอบเฉพาะ 20 ข้อ และข้อสอบร่วมกับชั้นประถมศึกษาปีที่ 5 จำนวน 5 ข้อ รวมเป็น 25 ข้อ มีค่าความเที่ยงเท่ากับ .89

2. แบบทดสอบที่ใช้สอบกับกลุ่มตัวอย่างสอบทานผล เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนของชั้นประถมศึกษาปีที่ 4, 5 และ 6 ทั้งสามฉบับมีข้อสอบร่วมภายในระหว่างชั้นแต่ละชั้น รวม 65 ข้อ

ผู้วิจัยได้รวมข้อสอบทั้ง 3 ฉบับมาเป็นฉบับเดียวกัน ประกอบด้วยข้อสอบเฉพาะสำหรับชั้นประถมศึกษาปีที่ 4 จำนวน 20 ข้อ ข้อสอบร่วมระหว่างชั้นประถมศึกษาปีที่ 4 กับชั้นประถมศึกษาปีที่ 5 จำนวน 5 ข้อ ข้อสอบเฉพาะสำหรับชั้นประถมศึกษาปีที่ 5 จำนวน 15 ข้อ ข้อสอบร่วมระหว่างชั้นประถมศึกษาปีที่ 5 กับชั้นประถมศึกษาปีที่ 6 จำนวน 5 ข้อ และข้อสอบเฉพาะสำหรับชั้นประถมศึกษาปีที่ 6 จำนวน 20 ข้อ รวมทั้งสิ้นเป็น 65 ข้อ มีค่าความเที่ยงเท่ากับ .83

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลด้วยตนเอง ในระหว่างวันที่ 15-19 กุมภาพันธ์ 2549 ได้ข้อมูลครบร้อยละ 100

การวิเคราะห์ข้อมูล

1. วิเคราะห์ค่าสถิติพื้นฐานของกลุ่มตัวอย่าง โดยใช้โปรแกรม SPSS
2. การเปรียบเทียบคะแนนแต่ละวิธี

2.1 วิธีเชิงเส้นตรง

2.1.1 คำนวณค่าความชัน (m) และจุดตัดแกน (c) ของสมการเส้นตรง ดังนี้

$$\mu_j = a\mu_i + b$$

$$\sigma_j = a\sigma_i$$

$$\text{ดังนั้น } m = \frac{\sigma_j}{\sigma_i} \text{ และ } c = \mu_j - \frac{\sigma_j}{\sigma_i} \mu_i$$

เมื่อ μ_j เป็น คะแนนเฉลี่ยของนักเรียนในระดับชั้น j

μ_i เป็น คะแนนเฉลี่ยของนักเรียนในระดับชั้น i

σ_j เป็น ส่วนเบี่ยงเบนมาตรฐานของคะแนนของนักเรียนในระดับชั้น j

σ_i เป็น ส่วนเบี่ยงเบนมาตรฐานของคะแนนของนักเรียนในระดับชั้น i

2.1.2 นำค่าที่คำนวณได้แทนลงในสมการเส้นตรง เพื่อใช้ปรับคะแนนจากแบบทดสอบในแต่ละระดับให้เป็นคะแนนสมมูลกัน (Kolen & Brennan, 1995) ดังนี้

$$x_{i*} = mx_i + c$$

เมื่อ x_{i*} เป็น คะแนนสมมูลของระดับชั้นที่ต้องการปรับเทียบคะแนน

m เป็น ค่าความชันของกราฟเส้นตรง

x_i เป็น คะแนนสอบของระดับชั้นที่ต้องการปรับเทียบคะแนน

c เป็น ค่าคงที่ของสมการ หรือจุดตัดแกนของกราฟเส้นตรง

2.1.3 สร้างตารางปรับเทียบคะแนนทั้ง 3 ระดับชั้น

2.2 วิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

2.2.1 คำนวณค่าความยาก ค่าอำนาจจำแนกของข้อสอบโดยใช้โปรแกรม BILOG-MG 3.0

2.2.2 คำนวณค่าความสามารถของผู้สอบ (θ) โดยใช้โปรแกรม BILOG-MG 3.0

2.2.3 ปรับค่าความสามารถของผู้สอบ (θ) ให้สมมูลกัน โดยใช้หลักการปรับเทียบเชิงเส้นตรง

(Hambleton & Swanminathan, 1985) ดังนี้

$$\theta_{ij} = \frac{a_i}{a_j} \theta_i + b_j - \frac{a_i}{a_j} b_i$$

เมื่อ θ_{ij} เป็น ค่าความสามารถที่ปรับแล้วของนักเรียนระดับชั้น i ที่ปรับมาอยู่ระดับชั้น j

θ_i เป็น ค่าความสามารถของนักเรียนระดับชั้น i

a_i เป็น ค่าอำนาจจำแนกเฉลี่ยของข้อสอบร่วมในระดับชั้น i

a_j เป็น ค่าอำนาจจำแนกเฉลี่ยของข้อสอบร่วมในระดับชั้น j

b_i เป็น ค่าความยากเฉลี่ยของข้อสอบร่วมในระดับชั้น i

b_j เป็น ค่าความยากเฉลี่ยของข้อสอบร่วมในระดับชั้น j

2.2.4 สร้างตารางปรับเทียบคะแนนทั้ง 3 ระดับชั้น โดยการแทนค่าในสมการที่ได้จากข้อ 3

3. คำนวณค่าความคลาดเคลื่อนของการปรับเทียบคะแนน โดยใช้สูตร

$$E = |x_i - x_i^*|$$

เมื่อ E เป็น ค่าความคลาดเคลื่อนของการปรับเทียบคะแนน

x_i เป็น คะแนนของแบบทดสอบชุดที่ i ที่ได้จากการสอบ

x_i^* เป็น คะแนนของแบบทดสอบชุดที่ i ที่ได้จากรายการปรับเทียบคะแนน

4. คำนวณค่าความเพี้ยนพองของการปรับเทียบคะแนน โดยตัดสินจากดัชนีความแตกต่าง

(Discrepancy Index: C) ที่ได้จากกลุ่มตัวอย่างสอบทานผล โดยใช้สูตรของปีเตอร์เสนและคณะ (1982) ดังนี้

$$C = \frac{\sum_{k=1}^n (x_{ik} - x_{jik})^2}{ns_i^2}$$

เมื่อ C เป็น ดัชนีความแตกต่าง

x_{ik} เป็น คะแนนจากแบบทดสอบฉบับที่ i ของนักเรียนคนที่ k

x_{jik} เป็น คะแนนจากแบบทดสอบฉบับที่ j ปรับเทียบไปเป็นคะแนนแบบทดสอบฉบับที่ i โดยใช้ตารางปรับเทียบคะแนนของนักเรียนคนที่ k

S_i^2 เป็น ความแปรปรวนของคะแนนจากแบบทดสอบฉบับที่ i

N เป็น จำนวนคนในกลุ่มตัวอย่างสอบทานผล

5. ประเมินดัชนีความแตกต่างตามหลักการประเมินของปีเตอร์เสนและคณะ (1982: 93-94) ซึ่งมีเกณฑ์การยอมรับดังนี้

ระดับน่าพอใจอย่างยิ่ง หมายถึง $C \leq (.05 S_x)^2$

ระดับน่าพอใจ หมายถึง $(.05 S_x)^2 < C \leq (.10 S_x)^2$

ระดับปานกลาง	หมายถึง	$(.10 S_x)^2 < C \leq (.15 S_x)^2$
ระดับไม่น่าพอใจ	หมายถึง	$(.15 S_x)^2 < C \leq (.20 S_x)^2$
ระดับไม่น่าพอใจอย่างยิ่ง	หมายถึง	$(.20 S_x)^2 < C$

6. ทดสอบความแตกต่างของค่าความคลาดเคลื่อนและค่าความเพียงพอของการเปรียบเทียบคะแนนระหว่างวิธีการเปรียบเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบ โดยใช้สถิติทดสอบที (t-test) ดังนี้ (Glass & Hopkins, 1996: 299)

$$t = \frac{\bar{X}_d}{S_{\bar{X}_d}}, \quad \bar{X}_d = \frac{\sum X_d}{n} \quad \text{และ} \quad S_{\bar{X}_d} = \sqrt{\frac{(\sum X_d^2 - n\bar{X}_d^2)/(n-1)}{n}}$$

เมื่อ X_d เป็น ความแตกต่างระหว่างคะแนนแต่ละคู่

n เป็น จำนวนคู่ของผลต่าง

ผลการวิจัย

1. ผลการเปรียบเทียบค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนน

ผู้วิจัยได้คำนวณค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนน ค่าเฉลี่ยของค่าความคลาดเคลื่อน และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนของแบบทดสอบทั้งฉบับแล้วนำมาเปรียบเทียบค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ โดยใช้สถิติทดสอบที (t-test) ดังมีรายละเอียดตามตารางที่ 3 และที่ 4

ตารางที่ 3 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน

ความคลาดเคลื่อน	วิธีเชิงเส้นตรง			วิธีทฤษฎีการตอบสนองข้อสอบ		
	n	M	SD	n	M	SD
E_{45}	300	0.736	0.217	300	1.615	1.040
E_{56}	300	0.815	0.315	300	1.668	1.232
รวมทั้งฉบับ	600	0.776	0.266	600	1.641	1.136

E_{45} แทน ค่าความคลาดเคลื่อนของคะแนนสมมูลของนักเรียนชั้นประถมศึกษาปีที่ 4

โดยเปรียบเทียบให้มีความเท่าเทียมกันกับนักเรียนชั้นประถมศึกษาปีที่ 5

E_{56} แทน ค่าความคลาดเคลื่อนของคะแนนสมมูลของนักเรียนชั้นประถมศึกษาปีที่ 5

โดยเปรียบเทียบให้มีความเท่าเทียมกันกับนักเรียนชั้นประถมศึกษาปีที่ 6

จากตารางที่ 3 แสดงว่า การปรับเทียบคะแนนวิธีเชิงเส้นตรงมีค่าเฉลี่ยของความคลาดเคลื่อนรวมทั้งฉบับเท่ากับ 0.776 และมีค่าส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 0.266 และ การปรับเทียบมาตราวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ มีค่าเฉลี่ยของความคลาดเคลื่อนรวมทั้งฉบับเท่ากับ 1.641 และมีค่าส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 1.136

ตารางที่ 4 เปรียบเทียบค่าความคลาดเคลื่อนของการปรับเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบ

วิธีการปรับเทียบคะแนน	n	M	SD	t
วิธีทฤษฎีการตอบสนองข้อสอบ	600	1.641	1.136	23.627**
วิธีเชิงเส้นตรง	600	0.776	0.266	

**p < .01

จากตารางที่ 4 แสดงว่า ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่วิธีเชิงเส้นตรงให้ค่าความคลาดเคลื่อนที่น้อยกว่าการปรับเทียบวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

2. ผลการวิเคราะห์และการเปรียบเทียบค่าความเพียงพอของการปรับเทียบคะแนน

2.1 การวิเคราะห์ค่าความเพียงพอของการปรับเทียบคะแนน

จากการนำแบบทดสอบวิชาคณิตศาสตร์ เรื่อง ทศนิยม ระดับชั้นประถมศึกษาปีที่ 4, 5 และ 6 ไปทดสอบกับกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 4, 5 และ 6 ระดับชั้นละ 120 คน โดยแยกแบบทดสอบออกเป็น 9 ฉบับ ดังนี้

1. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 4 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 4
2. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 4 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 5
3. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 4 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 6
4. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 5 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 4

5. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 5 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 5

6. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 5 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 6

7. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 6 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 4

8. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 6 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 5

9. แบบทดสอบของกลุ่มตัวอย่างสอบทานผลระดับชั้นประถมศึกษาปีที่ 6 จากการทำแบบทดสอบชั้นประถมศึกษาปีที่ 6

ผู้วิจัยได้นำคะแนนจากแบบทดสอบทั้ง 9 ฉบับนี้มาเปรียบเทียบมาตราทั้งวิธีเชิงเส้นตรงและวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ และหาค่าความเพี้ยนพองของการเปรียบเทียบคะแนน โดยตัดสินจากค่าดัชนีความแตกต่าง (C) แล้วประเมินค่าดัชนีความแตกต่าง ได้ผลการประเมินค่าดัชนีความแตกต่าง (C) ดังตารางที่ 5 และ 6

ตารางที่ 5 ค่าดัชนีความแตกต่าง (C) เมื่อเปรียบเทียบคะแนนสอบของกลุ่มตัวอย่างสอบทานผลจำนวน 120 คน ในแต่ละระดับชั้น โดยใช้วิธีเชิงเส้นตรง

กลุ่มตัวอย่าง สอบทานผล	การเปรียบเทียบ	$\sum(x_{ik} - x_{jik})^2$	SD ²	SD	C
ระดับชั้นป.4	ป.4 ปรึบให้เท่าเทียม ป.5	89.305	23.497	4.846	0.032 ⁽¹⁾
	ป.5 ปรึบให้เท่าเทียม ป.6	44.330	12.709	3.565	0.029 ⁽²⁾
ระดับชั้นป.5	ป.4 ปรึบให้เท่าเทียม ป.5	112.997	19.033	4.363	0.049 ⁽³⁾
	ป.5 ปรึบให้เท่าเทียม ป.6	96.153	29.099	5.394	0.028 ⁽⁴⁾
ระดับชั้นป.6	ป.4 ปรึบให้เท่าเทียม ป.5	133.640	17.586	4.194	0.063 ⁽⁵⁾
	ป.5 ปรึบให้เท่าเทียม ป.6	137.718	24.899	4.990	0.046 ⁽⁶⁾
					$M_C = 0.041$
					$SD_C = 0.014$

หมายเหตุ (1) ระดับน่าพอใจอย่างยิ่ง ($C \leq 0.059$) (2) ระดับน่าพอใจอย่างยิ่ง ($C \leq 0.032$)
 (3) ระดับน่าพอใจ ($0.048 < C \leq 0.190$) (4) ระดับน่าพอใจอย่างยิ่ง ($C \leq 0.073$)
 (5) ระดับน่าพอใจ ($0.044 < C \leq 0.176$) (6) ระดับน่าพอใจอย่างยิ่ง ($C \leq 0.062$)

จากตารางที่ 5 แสดงว่า ค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนจากคะแนนของกลุ่มตัวอย่างสอบทานผลโดยใช้วิธีเชิงเส้นตรง มีค่าเฉลี่ยเท่ากับ 0.041 และค่าส่วนเบี่ยงเบนมาตรฐานของดัชนีความแตกต่าง (C) มีค่าเท่ากับ 0.014 และเมื่อประเมินค่าดัชนีความแตกต่าง (C) พบว่า ส่วนใหญ่ให้ความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับน่าพอใจอย่างยิ่ง

ตารางที่ 6 ค่าดัชนีความแตกต่าง (C) เมื่อปรับเทียบคะแนนสอบของกลุ่มตัวอย่างสอบทานผลจำนวน 120 คน ในแต่ละระดับชั้น โดยใช้วิธีทฤษฎีการตอบสนองข้อสอบ

กลุ่มตัวอย่าง สอบทานผล	การปรับเทียบ	$\sum (x_{ik} - x_{jik})^2$	SD ²	SD	C
ระดับชั้นป.4	ป.4 ปรับให้เท่าเทียม ป.5	463.788	17.481	4.181	0.221 ⁽¹⁾
	ป.5 ปรับให้เท่าเทียม ป.6	237.139	6.972	2.641	0.283 ⁽²⁾
ระดับชั้นป.5	ป.4 ปรับให้เท่าเทียม ป.5	374.313	14.185	3.766	0.220 ⁽³⁾
	ป.5 ปรับให้เท่าเทียม ป.6	646.158	24.131	4.912	0.223 ⁽⁴⁾
ระดับชั้นป.6	ป.4 ปรับให้เท่าเทียม ป.5	350.541	13.405	3.661	0.218 ⁽⁵⁾
	ป.5 ปรับให้เท่าเทียม ป.6	545.306	19.911	4.462	0.228 ⁽⁶⁾
					$M_C = 0.232$
					$SD_C = 0.025$

หมายเหตุ (1) ระดับปานกลาง ($0.175 < C \leq 0.393$) (2) ระดับไม่น่าพอใจอย่างยิ่ง ($0.279 < C$)
 (3) ระดับปานกลาง ($0.142 < C \leq 0.319$) (4) ระดับน่าพอใจ ($0.060 < C \leq 0.241$)
 (5) ระดับปานกลาง ($0.134 < C \leq 0.302$) (6) ระดับปานกลาง ($0.199 < C \leq 0.448$)

จากตารางที่ 6 แสดงว่า ค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนจากคะแนนของกลุ่มตัวอย่างสอบทานผลโดยใช้วิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์มีค่าเฉลี่ยเท่ากับ 0.232 และค่าส่วนเบี่ยงเบนมาตรฐานของค่าดัชนีความแตกต่าง (C) มีค่าเท่ากับ 0.025 และเมื่อประเมินค่าดัชนีความแตกต่าง (C) พบว่า ส่วนใหญ่ให้ความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับปานกลาง

2.2 การเปรียบเทียบความเพียงพอของการปรับเทียบคะแนน

ผู้วิจัยได้หาค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนวิธีเส้นตรงและวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ในแต่ละระดับชั้น แล้วนำมาเปรียบเทียบค่าดัชนีความแตกต่าง (C) เพื่อประเมินความแตกต่างของความเพียงพอของการปรับเทียบคะแนนโดยใช้สถิติทดสอบที (t-test) ได้ผลดังตารางที่ 7

ตารางที่ 7 เปรียบเทียบค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบ

วิธีการปรับเทียบคะแนน	จำนวน C	M _C	SD _C	t
วิธีทฤษฎีการตอบสนองข้อสอบ	6	0.232	0.025	13.666**
วิธีเชิงเส้นตรง	6	0.041	0.014	

**p < .01

จากตารางที่ 7 แสดงว่า ค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่ค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนวิธีเชิงเส้นตรงน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แสดงว่า วิธีเชิงเส้นตรงมีความเพียงพอของการปรับเทียบมาตราอยู่ในระดับที่น่าพอใจมากกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

ผลการวิจัยสรุปได้ดังนี้

1. ผลการวิเคราะห์ความคลาดเคลื่อนของวิธีเชิงเส้นตรงมีความคลาดเคลื่อนเฉลี่ยเท่ากับ 0.78 และวิธีทฤษฎีการตอบสนองข้อสอบมีความคลาดเคลื่อนเฉลี่ยเท่ากับ 1.64 เมื่อเปรียบเทียบความแตกต่างของความคลาดเคลื่อนของการปรับเทียบคะแนนตามแนวตั้งระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ปรากฏว่า ความคลาดเคลื่อนของวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่ความคลาดเคลื่อนของวิธีเชิงเส้นตรงน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

2. ผลการวิเคราะห์ความเพียงพอของการปรับเทียบคะแนนตามแนวตั้ง โดยพิจารณาจากค่าดัชนีความแตกต่าง (C) พบว่า วิธีเชิงเส้นตรงมีค่าดัชนีความแตกต่างเฉลี่ยเท่ากับ 0.04 และวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์มีค่าดัชนีความแตกต่างเฉลี่ยเท่ากับ 0.23 เมื่อเปรียบเทียบความแตกต่างของค่าดัชนีความแตกต่างระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ปรากฏว่า ค่าดัชนีความแตกต่างของวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่ค่าดัชนีความแตกต่างของวิธีการปรับเทียบเชิงเส้นตรงน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ แสดงว่า วิธีเชิงเส้นตรงมีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจมากกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

การปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายในกับแบบวัดผลสัมฤทธิ์ทางการเรียน
วิธีเชิงเส้นตรงมีคุณภาพดีกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

การอภิปรายผลการวิจัย

ผลการวิจัยในครั้งนี้ มีประเด็นสำคัญที่นำมาอภิปราย ดังต่อไปนี้

1. การเปรียบเทียบความคลาดเคลื่อนของวิธีการปรับเทียบคะแนน วิธีเชิงเส้นตรงเป็นวิธีที่มีความคลาดเคลื่อนน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ซึ่งไม่สอดคล้องกับสมมติฐาน ทั้งนี้ อาจเนื่องมาจากการวิจัยนี้ได้ใช้การปรับเทียบคะแนนตามแนวตั้งวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ โดยไม่ได้ใช้ค่าการเดาเข้ามาเกี่ยวข้อง ซึ่งในทางปฏิบัติจริง แบบทดสอบที่นำไปสอบกับผู้สอบนั้นย่อมมีค่าการเดาเกิดขึ้น จึงอาจเป็นสาเหตุทำให้การปรับเทียบคะแนนตามแนวตั้งวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์มีความคลาดเคลื่อนมาก จากการศึกษาของ ลอร์ด (1980) ได้สรุปว่า การปรับเทียบคะแนนตามแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ หากใช้การปรับเทียบด้วยคะแนนจริง จะไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ถึงแม้จะเป็นการปรับเทียบโดยใช้คะแนนจริง แต่เป็นค่าที่ประมาณได้จากสูตรการคำนวณ ดังนั้นจึงยังคงมีความคลาดเคลื่อนอยู่ และจากการศึกษาของ โคลน (1981) ได้สรุปว่า การปรับเทียบคะแนนโดยใช้วิธีทฤษฎีการตอบสนองข้อสอบแบบหนึ่งและสองพารามิเตอร์ยังเป็นวิธีที่ไม่ถูกต้อง เนื่องมาจากวิธีทฤษฎีการตอบสนองข้อสอบแบบหนึ่งและสองพารามิเตอร์ไม่ได้คำนึงถึงค่าการเดา แต่การทำข้อสอบของผู้สอบมีค่าการเดาเกิดขึ้นด้วย จึงทำให้ไม่สามารถปรับเทียบคะแนนจากแบบทดสอบที่มีค่าการเดาเกิดขึ้นได้ อีกสาเหตุหนึ่งที่ทำให้วิธีเชิงเส้นตรงมีความคลาดเคลื่อนน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ คือ การวิจัยนี้ใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ซึ่งมีค่าความเที่ยงอยู่ในเกณฑ์สูง วัดในเนื้อหาและจุดประสงค์ชุดเดียวกัน ทำให้การแจกแจงคะแนนของแบบทดสอบมีลักษณะคล้ายคลึงกัน ซึ่งเป็นไปตามนิยามการปรับเทียบคะแนนเชิงเส้นตรง ผลการปรับเทียบคะแนนวิธีเชิงเส้นตรงจึงให้ผลเป็นที่น่าพอใจ

2. การเปรียบเทียบความเพียงพอของการปรับเทียบคะแนนตามแนวตั้ง โดยพิจารณาจากค่าดัชนีความแตกต่าง (C) วิธีเชิงเส้นตรงเป็นวิธีที่มีค่าดัชนีความแตกต่างน้อยกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ทำให้วิธีเชิงเส้นตรงมีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจมากกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ ซึ่งไม่สอดคล้องกับสมมติฐาน อาจเป็นผลมาจากการปรับเทียบคะแนนในงานวิจัยนี้ ใช้กับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน การทดสอบวัดผลสัมฤทธิ์ทางการเรียนจะวัดตามหลักสูตร ซึ่งมีขอบเขตเนื้อหาที่จำกัด ทำให้แบบทดสอบในระดับชั้นที่สูงกว่าประกอบด้วยข้อสอบที่มีลักษณะต่าง ๆ ที่ใกล้เคียงกับแบบทดสอบที่สร้างขึ้นในระดับชั้นที่ต่ำกว่าเมื่อ แบบทดสอบวัดเนื้อหาเดียวกันแต่มี

ความยากต่างกัน วิธีการปรับเทียบคะแนนเชิงเส้นตรงจึงให้ผลดีกว่าวิธีอื่น (Kolen & Whitney, 1982) ซึ่งสอดคล้องกับผลการศึกษาของภาวิณี ศรีสุขวัฒนานันท์ (2530) ที่พบว่าในกรณีที่ใช้แบบทดสอบวัดผลสัมฤทธิ์วิธีเชิงเส้นตรงให้ผลเป็นที่น่าพอใจมากที่สุด และเมื่อทดสอบความแตกต่างของค่าดัชนีความแตกต่างเป็นรายคู่แล้ว พบว่าวิธีเชิงเส้นตรงมีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจ มากกว่าวิธีทฤษฎีการตอบสนองข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และจากการศึกษาของรุ่งนภา จันทรา (2540) ที่พบว่า การปรับเทียบคะแนนโดยวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ (ไม่มีค่าการเดา) จะไม่สามารถอธิบายผลการปรับเทียบคะแนนได้ จึงควรใช้วิธีการปรับเทียบคะแนนเชิงเส้นตรงจะได้ผลที่ดีกว่า

ข้อเสนอแนะ

การนำผลการวิจัยไปใช้

ในการปรับเทียบคะแนนตามแนวดิ่ง กรณีที่เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ควรใช้วิธีเชิงเส้นตรง เนื่องจากเป็นวิธีที่มีค่าความคลาดเคลื่อนต่ำมีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจ อีกทั้งยังเป็นวิธีที่มีความเหมาะสมและถูกต้องตามสถานการณ์ในการปรับเทียบคะแนนมากกว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์

การทำวิจัยต่อไป

1. ควรมีการศึกษาเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวดิ่งในวิธีอื่นๆ เช่น วิธีอีควิเปอร์เซ็นไทล์ วิธีใช้สมการถดถอย เป็นต้น
2. ควรเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวดิ่งระหว่างการใช้แบบทดสอบร่วมภายในกับแบบทดสอบร่วมภายนอก

เอกสารอ้างอิง

- เกตุรินทร์ หมัดอะดัม. (2545). การเปรียบเทียบคุณภาพของการเทียบมาตรฐานแนวดิ่งระหว่างรูปแบบทฤษฎีการตอบสนองที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยทักษิณ.
- พรพิมล นาคเวช. (2537). การศึกษาคุณภาพของการเทียบมาตรฐานแนวดิ่งโดยใช้ทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

- ภาวิณี ศรีสุขวัฒนานนท์. (2530). *การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่ต่างกันเมื่อแบบสอบร่วมมีความยาวต่างกัน*. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- รุ่งนภา จันทรา. (2540). *การเปรียบเทียบความคลาดเคลื่อนและความเพียงพอของการเทียบมาตราระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์*. วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต, สาขาวิชาการวัดผลและวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยสงขลานครินทร์.
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, N.J.: Educational Testing Service.
- Ayerve, R. I. (1992). The effectiveness of equipercentile method and IRT three – parameter Model on vertical equating under varying conditions of sample size, test length, and anchor test length: a simulation study. *Dissertation Abstracts International*, 34(5), 143-168.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical method in education and psychology*. 3rd.th. N. J.: Prentice – Hall.
- Hambleton, R. K., & Swanminathan, H. (1985). *Item response theory: Principle and application*. Boston: Kluwer-Nijhaff publishing.
- Harris, D. J. (1991). A comparison of Angoff's Design I and Design II for vertical equating using tradition and IRT methodology. *Journal of Educational Measurement*, 28(74), 221 -234.
- Hill, J. R., & Subhiyah, R. G. (1988). Equating minimum-competency tests: Comparisons of methods. *Journal of Educational Measurement*, 25(3), 221-231.
- Johanson, G. A. (1988). A study of item response theory equating with an anchor test design. *Dissertation Abstracts International*, 5(63), 11-12.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18(Spring 1981), 1-11.
- Kolen, M. J., & Whitney, D. R. (1982). Comparison of four procedures for equating the tests of general educational development. *Journal of Educational Measurement*, 19(Winter 1982), 279 – 292.
- Kolen, M. J., & Brennan, R. L. (1987). A reply to Angoff. *Applied Psychological Measurement*, 11(3), 301 - 306.
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer -Verlag.
- Leung, Shing On. (2003). *A practical use of vertical equating by combining IRT equating and linear equating*. *Practical Assessment, Research & Evaluation*, 8(23). Retrieved August 17, 2004 from <http://PAREonline.net/getvn.asp?v=8&n=23>.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

- Petersen, N. S., Marco, C. L., & Stewart, E. E. (1982). A test of the adequacy of linear score equating methods. In P.W. Holland, and D.B. Rubin (ed), *Test Equating*. New York: Academic Press.
- Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8, 137 – 156.