ABSTRACT

The purpose of this research was to study the use of GRM, GPCM and logistic models to compare test information functions of tests employing dichotomous (0, 1) and polytomous (1, 2, 3, 4) scoring methods. Data obtained from dichotomous and polytomous scoring methods were analyzed based on 1, 2, 3 parameter logistic models, and based on GRM and GPCM, respectively. Two data sets were conducted in this study: (1) general affective scale data of 6,300 Prathom Suksa 6 students from the Office of Educational Assessment and Testing Service, Department of Curriculum and Instruction Development, Ministry of Education (2) mathematics achievement data of 470 Prathom Suksa 6 students collected by the researcher. Detections of person and item fits were performed through RSM using BIGSTEPS.
program. Test information functions were determined by MULTILOG and PARSCALE programs. Confirmatory factor analysis was used to examine construct validity of general affective scale through LISREL program.

Results showed that polytomous scoring method based on GRM provided higher test information function than dichotomous scoring method. Nevertheless, there was no clear evidence whether the polytomous scoring method based on GPCM yielded higher test information function than dichotomous scoring method.

บทัศน์ย่อ

การวิจัยนี้มีจุดมุ่งหมายเพื่อศึกษาการใช้ GRM, GPCM และโมเดลโลจิสติกในการเปรียบเทียบค่าพังก์ชันสารสนเทศของแบบวัดที่มีการตรวจให้คะแนนแบบทรัพยากร (0, 1) และแบบพฤติกรรม (1, 2, 3, 4) ข้อมูลจากการตรวจให้คะแนนแบบทรัพยากรวิเคราะห์ตามโมเดลโลจิสติก, 2 และ 3 พรามเทอร์ ส่วนการตรวจให้คะแนนแบบพฤติกรรมวิเคราะห์ตาม GRM และ GPCM การวิจัยครั้งนี้ทำการศึกษาเก็บข้อมูล 2 ชุด (1) ข้อมูลผลการวัดคุณลักษณะทั่วไป ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 6,300 คน จากสำนักงานเขตสอบทางการศึกษา กรมวิชาการการตรวจศึกษา (2) ข้อมูลผลผลิตบุตร์ทางคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 470 คน ที่ผู้วิจัยเก็บเอง ทำการตรวจสอบความสมบูรณ์ของข้อมูลโดยวิเคราะห์ความเหมาะสมของผู้ตอบและข้อกระทำตาม RSM ด้วยโปรแกรม BIGSTEPS และวิเคราะห์เพื่อหาค่าพังก์ชันสารสนเทศของแบบวัดโดยใช้โปรแกรม MULTILOG และ PARSCALE จากนั้นตรวจสอบความตรงชี้โครงสร้างของแบบวัดคุณลักษณะทั่วไปโดยวิเคราะห์โครงสร้างเชิงอนุมาน ด้วยโปรแกรม LISREL

ผลการวิจัยพบว่าการตรวจให้คะแนนแบบพฤติกรรมเมื่อวิเคราะห์ตาม GRM ให้ค่าพังก์ชันสารสนเทศสูงกว่าการตรวจให้คะแนนแบบทรัพยากร สำหรับการตรวจให้คะแนนแบบพฤติกรรมวิเคราะห์ตาม GPCM
และแบบทดสอบที่ยังไม่สามารถมาตรฐานได้ว่าที่ได้ให้คำพังก์ขึ้นสารสนเทศสูงกว่ากัน

ความเป็นมาและความสำคัญของปัญหา

แบบวัด (tests) เป็นเครื่องมือวัดผลทางการศึกษาประเภทหนึ่งที่นิยมใช้กันมากเนื่องจากมีความสะดวกในการนำไปใช้ คือช่วยกำกับระดับของคุณลักษณะออกมาในรูปคะแนน ซึ่งเป็นข้อมูลเชิงปริมาณ แบบวัดดังกล่าวแบ่งได้เป็น 2 ชนิดตามลักษณะการนำไปใช้ คือ (1) แบบวัดคุณลักษณะ โดยทำไปมักเป็นมาตรฐานค่า (rating scale) (2) แบบสอบวัดผลสมการทางการเรียน (achievement tests) แบบวัดทั้ง 2 ชนิดดังกล่าวมีวิธีให้คะแนน 2 แบบ คือ การให้คะแนนแบบทริวิวติกา (0, 1) และการให้คะแนนแบบพหุวิภาคอย่างไรก็ตามในการนำแบบวัดคุณลักษณะที่เป็นมาตรฐานค่าแบบตัวเลือกบังคับตอบ (force - choice ratings) และแบบสอบวัดผลผลิตที่ใช้พบว่ามีการตรวจสอบให้คะแนน 2 แบบ คือแบบทริวิวติกา โดยให้ตัวเลือกที่ถูกต้องหรือเหมาะสมที่สุดเป็น 1 คะแนน ส่วนตัวเลือกอื่น ๆ เป็น 0 คะแนน (วาร์นีเยิมาร์มม, 2536; บูรินทร์, บุญรอด, 2536) ซึ่งวิธีนี้ทำได้สะดวกและรวดเร็ว และถือว่าเน้นที่การให้คะแนนแบบพหุวิภาค โดยกำหนดค่าคะแนนให้กับตัวเลือกที่ถูกต้องต่างกันตามระดับของคุณลักษณะ ซึ่งการใช้ทริวิวติกาในการวิจัยทั่วไป (สุพัตรา เกยนอุดม, 2536; ฤทธิ์ ชนะพงศ์, 2534; บุญทอง บุญทกวี, 2534)

การวัดที่ตรงและเที่ยงก็อาจจะพิจารณาถึงวิธีการให้คะแนนที่เหมาะสมแล้ว ภาระ

การวิเคราะห์เพื่อประมาณค่าความสามารถของผู้สอบ และตรวจสอบคุณภาพของข้อสอบก็เป็น

ภาระที่สำคัญของผู้ที่จะทำให้กลุ่มสเปนภัยกับบุคคลและข้อรองที่มีความตรงและเที่ยง การวิเคราะห์ตั้งกล่าวจานกับเป็น 2 แนว คือ (1) การวิเคราะห์ตามแนวทฤษฎีการตอบแบบตระกูล (classical test theory : CTT) ภาระนี้มีข้อจัดเกณฑ์จากคำศัพท์ที่ได้มักแปล

เปรียบกับผลลัพธ์ด้วยอย่าง (2) การวิเคราะห์ตามแนวทฤษฎีการตอบแบบช่องสเปน (item response theory : IRT) ภาระนี้จะให้คำวิเคราะห์ที่สามารถอ้างอิงได้กับกลุ่มตัวอย่างทั่วไป ซึ่งจานกับการวิเคราะห์เป็น 2 แนวทางตามวิธีการให้คะแนน คือ (1) การวิเคราะห์ตาม Dichotomous IRT Model ซึ่งใช้กับแบบวัดที่ตรงให้คะแนนแบบบริวาร ได้แก่การ

วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 ภาระมิเตอร์ เป็นต้น (2) การวิเคราะห์ตาม Polytomous IRT Model ซึ่งใช้กับแบบวัดที่ตรงให้คะแนนแบบบริวาร มีตัวกำหนดหลายโมเดล (Donoghue, 1994; Muraki, 1993) เช่น Graded Response Model (GRM), Nominal Response Model (NRM), Rating Scale Model (RSM), Partial Credit Model (PCM),

Successive Interval Model (SIM) และ Generalized Partial Credit Model (GPCM)

สำหรับโมเดลที่มีความยืดหยุ่นและมีการนำไปใช้กันมากในปัจจุบัน คือ GRM ซึ่งพัฒนามา

จากโมเดลโลจิสติก 2 ภาระมิเตอร์ และ GPCM ซึ่งพัฒนามาจากโมเดลรายชิ้น 1 ภาระภาระมิเตอร์ (Donoghue, 1994; Muraki, 1993) จากวิธีการวิเคราะห์ตั้งกล่าวจึงเป็นประเด็นสำหรับการ

วิจัยตามความวิวัฒนาให้คะแนนตั้ง 2 วิธี คือ แบบบริวารและแบบบริวาร ที่วิเคราะห์ตาม

โมเดลโลจิสติก GRM และ GPCM เมื่อใช้กับแบบวัดคุณลักษณะและแบบทดสอบ ผลลัพธ์จะให้ค่าพักรักษาสารสนเทศต่างกันหรือไม่ ถ้าต่างกัน วิธีใดจะให้ค่าพักรักษาสารสนเทศ

สูงกว่ากัน

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบค่าพักรักษาสารสนเทศของแบบวัดคุณลักษณะและแบบทดสอบ (TIF)

สำหรับการตรวจสอบให้คะแนนแบบบริวารและแบบบริวารเมื่อวิเคราะห์ตาม GRM และ

GPCM

2. เพื่อเปรียบเทียบอัตราสารสนเทศเฉลี่ย (ratio of average information : RAI) ของแบบวัดคุณลักษณะและแบบทดสอบสำหรับการตรวจสอบให้คะแนนแบบบริวารและแบบ

บริวารเมื่อวิเคราะห์ตาม GRM และ GPCM
3. เพื่อตรวจสอบความสอดคล้องของคำพิพักษ์ขั้นสารสนเทศของแบบวัดคุณลักษณะและแบบสอบถามระหว่างการวิเคราะห์ตาม GRM และ GPCM

สมมติฐานของการวิจัย

1. ผลการวิเคราะห์การตรวจให้คะแนนทั้ง 2 วิธี เมื่อวิเคราะห์ตาม GRM และ GPCM ทั้ง 2 ไม่ได้ให้ผลที่สอดคล้องกัน คือ การตรวจให้คะแนนแบบพฤติกรรมที่วิเคราะห์ตาม GRM และ GPCM จะให้คำพิพักษ์ขั้นสารสนเทศ (TIF) สูงกว่าการตรวจให้คะแนนแบบพฤติกรรมที่วิเคราะห์ตามโมเดลจิตสิก

2. คำสารสนเทศเดียว (RAI) ของแบบวัดที่ใช้วิธีการตรวจให้คะแนนแบบพฤติภาพเมื่อวิเคราะห์ตาม GPCM จะให้คำพิพักษ์ขั้นสารสนเทศสูงกว่าการวิเคราะห์ตาม GRM

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่าง

ประชากรในการศึกษาครั้งนี้แบ่งเป็น 2 กลุ่ม คือ (1) นักเรียนชั้นประถมศึกษาปีที่ 6 ของทุกเขตการศึกษาที่เรียนในปีการศึกษา 2537 และเข้ารับการทดสอบโดยการใช้แบบวัดคุณลักษณะของสั่นภายนอก กรมวิชาการ รวมทั้งสิ้น จำนวน 102,117 คน (2) นักเรียนชั้นประถมศึกษาปีที่ 6 ใส่เสื้อผ้าสีบานทำการประเมินคุณสมบัติอย่างเพื่อประโยชน์ศูนย์ข้อมูลที่กล่าวเรียนในภาคเรียนที่ 2 ปีการศึกษา 2538 จำนวน 1,567 คน

กลุ่มตัวอย่างที่ศึกษามี 2 กลุ่ม คือ (1) นักเรียนที่เข้ารับการทดสอบโดยสำนักงานทดสอบทางการศึกษา กรมวิชาการ จำนวน 6,300 คน (2) นักเรียนในกลุ่มที่สำนักงานการประเมินคุณสมบัติอย่างเพื่อประโยชน์ศูนย์ข้อมูล จำนวน 470 คน

ตัวแปรที่ศึกษา

1. ตัวแปรตัวแปร

1.1 วิธีการตรวจให้คะแนนแบ่งเป็น 2 วิธี คือ การตรวจให้คะแนนแบบพฤติภาพและการตรวจให้คะแนนแบบพฤติภาพ

1.2 ไม่ได้ใช้วิเคราะห์เปรียบเทียบในการศึกษารั้งนี้ 3 ไม่ได้ คือ ไม่ได้การวิเคราะห์ GRM, GPCM และโมเดลจิตสิก
2. ค่าพักรักษาสุขภาพ
   2.1 อัตราส่วนสารสนเทศเฉพาะ (RAI)
   2.2 ความสอดคล้องของค่าพักรักษาสุขภาพของแบบวัดระหว่างการวิเคราะห์ตาม GRM และ GPCM

แบบวัดและแบบสอบ

เลือกเมื่อที่ใช้สำหรับเก็บรวบรวมข้อมูลในการศึกษาวันนี้มี 2 ชุด คือ

1. แบบวัดคุณลักษณะทางจิตใจ ที่พัฒนาขึ้นโดยนักทักษะลบ กรมวิชาการ กระทรวงศึกษาธิการ ซึ่งมีลักษณะเป็นมาตรประมาณค่าแบบสอบถามนี้มีคู่เปรียบเทียบด้วย มี 4 ตัวเลือก กำหนดให้คะแนนตามระดับของคุณลักษณะด้านเจริญพิสัย (affective domain) เป็น 1, 2, 3 และ 4 คะแนน จากตัวไปสู่ตามลำดับ

2. แบบสอบคณิตศาสตร์ ใช้วิเคราะห์สถิติปัญหาคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 พัฒนาขึ้นโดยวิศวฯ เลยส์มบ (2539) แบบสอบตัวกล่าวมี 6 ข้อ แต่ละข้อแบ่งเป็น 3 ตอน คือ ตอนที่ 1 วัดความเข้าใจวิชาการแก้โจทย์ปัญหา ตอนที่ 2 วัดการคิดคำนวณ และตอนที่ 3 วัดความสามารถในการวิเคราะห์ตามสภาพที่เป็นจริง โดยทั้ง 3 ตอนมีสถานการณ์และตัวเลขเหมือนกันหมด การตรวจให้คะแนนแบบพูดวิเคราะห์แบบสอบคณิตศาสตร์ในแต่ละข้อของแบบสอบถ้า 3 ตอน กำหนดเป็น 4 ลำดับขึ้น คือ นักเรียนที่ไม่ตอบหรือตอบไม่ถูกเลยทั้ง 3 ตอน ให้ 1 คะแนน นักเรียนที่ตอบถูกเพียง 1 ตอน ให้ 2 คะแนน นักเรียนที่ตอบถูก 2 ตอน ให้ 3 คะแนน และนักเรียนที่ตอบถูกทั้ง 3 ตอน ให้ 4 คะแนน

การวิเคราะห์เพื่อคัดเลือกผู้ตอบและข้อกระทำที่เหมาะสม

ในการวิจัยครั้งนี้ ได้ดำเนินการศึกษาจากข้อมูล 2 ชุด ข้อมูลชุดที่ 1 เป็นข้อมูลจาก การใช้แบบวัดทักษะและผู้ตอบทักษะ สำนักข้อมูลชุดที่ 2 เป็นข้อมูลจากผู้ตอบแบบเหมาะสม (person-fit) และข้อกระทำที่เหมาะสม (item-fit) ตาม Rating Scale Model (RSM) คัดเลือกโดยใช้โปรแกรม BIGSTEPS และหลังจากคัดเลือกข้อกระทำที่เหมาะสมแล้วพบว่าแบบวัดคุณลักษณะ มีจำนวนข้อกระทำต่ำ จึงได้ทำการตรวจสอบความตรงเชิงโครงสร้างโดยวิธีวิเคราะห์เชิงยืนยัน (confirmatory factor analysis : CFA) โดยใช้โปรแกรม LISREL หลังจากการวิเคราะห์พบว่าโครงสร้างของแบบวัดคุณลักษณะในข้อมูลชุดที่ 1
การวิเคราะห์ข้อมูล

d้านการวิเคราะห์ข้อมูลเบื้องต้นของคะแนนเกณฑ์ จากการใช้แบบวัดคุณลักษณะทั่วไป และแบบสอบถามคณิตศาสตร์ วิเคราะห์เพื่อคัดเลือกผู้ตอบที่เหมะสม (person fit) และ ข้อกระหะที่เหมะสม (item fit) โดยใช้โปรแกรมสำเร็จรูป BIGSTEPS วิเคราะห์องค์ประกอบจำลองยืนยัน (confirmatory factor analysis : CFA) เพื่อตรวจสอบความตรงชิ้นโครงสร้าง (construct validity) ของแบบวัดคุณลักษณะจากข้อมูลสูตรที่ 1 และ 2 ในแบบวัดย่อยที่ 1 และ 2 ตามลำดับโดยใช้สถิติทดสอบไซ-สแควร์ (chi-square) ด้วยโปรแกรมสำเร็จรูป LISREL วิเคราะห์หาค่าพหุจำนวนที่ออกของแบบวัดคุณลักษณะและแบบสอบถามจากข้อมูลสูตรที่ 1 และ 2 ด้วย GRM, GPCM และโมเดลโลจิสติก ด้วยโปรแกรม MULTILOG และ PARSCALE คำนวณค่าสารสนเทศเฉลี่ย (average information : AI) และอัตราส่วนสารสนเทศเฉลี่ย (RAI) โดยคำนวณแม่

สรุปผลการวิจัย

1. ค่าพหุจำนวนสารสนเทศ (TIF) ของแบบวัดคุณลักษณะและแบบสอบถามคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบวัดบริบทและการแบบพหุบริบท.

การวิเคราะห์ข้อมูลจากแบบวัดคุณลักษณะชุดที่ 1 และ 2 ตาม GRM ให้ผลการวิเคราะห์ที่สมดุลสมบัติ คือ ในช่วง 0 ถึง -2 ถึง 0 การตรวจให้คะแนนแบบพหุบริบทที่วิเคราะห์ตาม GRM ให้ค่าพหุจำนวนที่ออกกับการตรวจให้คะแนนแบบวัดบริบท ที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่ในช่วง 0 ถึง -1.5 การตรวจให้คะแนนแบบวัดบริบทที่วิเคราะห์ตามโมเดลโลจิสติก 3 พารามิเตอร์ ให้ค่าพหุจำนวนที่ออกกับการตรวจให้คะแนนแบบพหุบริบทที่วิเคราะห์ตาม GRM การวิเคราะห์ข้อมูลแบบสอบถามคณิตศาสตร์ชุดที่ 1 และ 2 ได้ผลที่สมดุลสมบัติ คือ ในช่วง 0 ถึง -2 ถึง -0.5 การตรวจให้คะแนนแบบพหุบริบทที่วิเคราะห์ตาม GRM ให้ค่าพหุจำนวนที่ออกกับการตรวจให้คะแนนแบบวัดบริบทที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่ในช่วง 0 ถึง -0.5 การตรวจให้คะแนนแบบวัดบริบทที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ ให้ค่าพหุจำนวนที่ออกกับการตรวจให้คะแนนแบบพหุบริบทที่วิเคราะห์ตาม GRM นอกจากนี้ยังพบว่าข้อมูลจากแบบวัดคุณลักษณะสูตรที่ 1 และ 2 ที่ใช้วิเคราะห์ให้คะแนนแบบวัดบริบทให้ผลที่สมดุลสมบัติ โดยในช่วง 0 ถึง -2 ถึง -0.5 การวิเคราะห์ตามโมเดลโลจิสติก 1 พารามิเตอร์ ให้ค่าพหุจำนวนที่ออกกับการวิเคราะห์ตาม...
โมเดลโลจิสติก 2 และ 3 พารามิเตอร์ตามลำดับ

2. การเปรียบเทียบอัตราส่วนสารสนเทศเฉลี่ยของแบบวัดและแบบสอบถาม สิ่งต่างๆ วิเคราะห์ได้คะแนนแบบที่ระหว่างตัววิเคราะห์ตาม GRM และ GPCM

การเปรียบเทียบอัตราส่วนสารสนเทศเฉลี่ยจากข้อมูลแบบวัดคู่คุณลักษณะชุดที่ 1 และ 2 ให้ผลการวิเคราะห์ที่สอดคล้องกัน คือ การตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GRM มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ และพบว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GPCM การวิเคราะห์ดังกล่าว เมื่อใช้กับแบบสอบถามคิดศาสตร์สังเขปข้อมูลสุ่มชุดที่ 1 และ 2 ก็ให้ผลที่สอดคล้องกัน คือ การตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GRM และ GPCM มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ ส่วนการเปรียบเทียบผลการตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GRM มีประสิทธิภาพสัมพันธ์สูงกว่าการวิเคราะห์ตาม GPCM เล็กน้อย

3. การพิจารณาความสอดคล้องของผลการตรวจสอบได้คะแนนแบบที่วิเคราะห์และแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GRM และ GPCM

ผลการวิเคราะห์ตาม GRM และ GPCM ที่สอดคล้องกัน คือ ผลจากการน้ำไปใช้กับแบบวัดคิดศาสตร์โดยพบว่าในข้อมูลแบบสอบถามคิดศาสตร์ชุดที่ 1 และ 2 ที่วิเคราะห์ตาม GRM และ GPCM มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่ผลที่จัดมาดังกล่าว คือ การน้ำไปใช้กับแบบวัดคู่คุณลักษณะ โดยพบว่าการตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GRM มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์ แต่การตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่วิเคราะห์ตาม GPCM มีประสิทธิภาพสัมพันธ์สูงกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 พารามิเตอร์

อภิปรายผลการวิจัย

ผลจากการศึกษาข้อสรุปได้ว่าการตรวจสอบได้คะแนนแบบพหุภาคภิวัติการที่เหมาะสมกว่าการตรวจสอบได้คะแนนแบบที่วิเคราะห์ โดยเฉพาะการตรวจสอบได้คะแนนที่ต่างกันมากเมื่อวิเคราะห์ตาม
GRM มีความเหมาะสมสำหรับการวิเคราะห์ตาม GPCM และโมเดลโลจิสติก 1, 2 และ 3
เพราะมีดีกรีที่สอดคล้องกับผลงานวิจัยของโดนิกว์ (Donoghue, 1994) และชาเมจิมา
(Samejima, 1976 cited by Donoghue, 1994) ที่ศึกษาปริมาณแบบฟังก์ชันสถานะหนังสือ
ของแบบสอบถามประเภทเขียนตอบ โดยใช้วิธีตรวจให้คะแนนแบบพฤติภาพและแบบทริปวิทยา
และพบว่าการตรวจให้คะแนนแบบพฤติภาพให้คำฟังก์ชันสถานะสูงกว่าการตรวจให้คะแนน
แบบทริปวิทยา นอกจากนี้ บอก ขั้มเลี้ยน และทิสเซ็น (Bock, Ssymson and Tissen cited
by Drasgow, F, 1995) ยังได้แนะนำให้ใช้การตรวจให้คะแนนแบบพฤติภาพกับแบบวิทยา
จิตวิทยาแม้กระนั้นการตรวจให้คะแนนแบบทริปวิทยา เพราะคะแนนที่ได้สำหรับสถานะมากกว่า
การใช้การตรวจให้คะแนนแบบทริปวิทยา

ผลการวิจัยครั้งนี้ไม่เป็นไปตามสมมุติฐานข้อ 2 ซึ่งการวิเคราะห์ตาม GRM และ
GPCM ให้ผลไม่สอดคล้องกันเมื่อใช้กับเครื่องมือที่เป็นแบบวัดคุณลักษณะทั่วไป โดยพบว่า
การตรวจให้คะแนนแบบพฤติภาพที่วิเคราะห์ตาม GRM ได้คำสารสนเทศเฉลี่ยสูงกว่าการ
ตรวจให้คะแนนแบบทริปวิทยาที่วิเคราะห์ตาม GPCM พบว่าได้คำสารสนเทศเฉลี่ย
ต่ำกว่าการตรวจให้คะแนนแบบทริปวิทยาที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 เพราะมีดีกรี
เมื่อพิจารณาจากงานวิจัยที่ผ่านมา (Donoghue, 1994 : Samejima, 1976 cited by
Donoghue, 1994) พบว่างานวิจัยล่าสุดล่าสุดใช้เครื่องมือที่เป็นแบบสอบถามประเภทวัดผล
สมัคร_rngษ์เช่น การศึกษาของโดนิกว์ (Donoghue, 1994) ใช้เครื่องมือที่เป็นแบบสอบถาม
ความสามารถในการอ่าน (reading assessment) วิเคราะห์ด้วย GPCM ได้ข้อตัวแปรที่
สอดคล้องกันว่าการตรวจให้คะแนนแบบพฤติภาพให้คำฟังก์ชันสถานะสูงกว่าการตรวจให้
คะแนนแบบทริปวิทยา การที่ผลการวิจัยครั้งนี้ไม่สอดคล้องกับสมมุติฐานล่าสุด คือ เพราะ
ลักษณะของเครื่องมือที่ใช้มีอิทธิพลต่อการวิเคราะห์ด้วย GPCM เพราะเมื่อพิจารณาจากผล
การใช้กับแบบสอบถามคินิติศาสตร์ การวิเคราะห์ตาม GPCM ได้คำฟังก์ชันสถานะสูงกว่า
การตรวจให้คะแนนแบบทริปวิทยาที่วิเคราะห์ตามโมเดลโลจิสติก ทั้งนี้เพราะนักเรียนที่ตอบ
แบบวัดคุณลักษณะมีการตอบแบบหลอก (faking) สูงกว่าการตอบแบบทาง (guessing) ใน
การทำแบบสอบถามวัดคุณลักษณะที่ใช้ในศึกษาครั้งนี้เป็นข้อมำะมากมายค่านักเรียนตัวเลือก
นั้นตอบ 4 ตัวเลือก ส่วนแบบสอบถามคินิติศาสตร์ประกอบด้วยข้อสอบถามคินิติศาสตร์ตัวเลือก 5 ตัว
เลือก โดยรวมตัวเลือกปลายเป็นตัวด้วย ทำให้เกิดข้อความที่คาดการณ์จากตัวแบบวัด
คุณลักษณะ
ผลการศึกษาที่ไม่เป็นไปตามสมมติฐานอีกประมาณหนึ่ง คือ การตรวจให้คะแนนแบบพื้นฐานที่ใช้แบบวัดคุณลักษณะและแบบสอบคณิตศาสตร์ เมื่อวิเคราะห์ตาม GRM ให้ค่าสารสนเทศเพิ่มขึ้นสูงกว่าการวิเคราะห์ตาม GPCM แสดงให้เห็นว่าการตรวจให้คะแนนแบบพื้นฐานที่วิเคราะห์ตาม GRM มีความเหมาะสมกว่าการวิเคราะห์ตาม GPCM ซึ่งจะเห็นได้จากการวิจัยของ อ. อะยาล่า De-Ayala (1992) ที่ศึกษาเปรียบเทียบผลการวิเคราะห์การตรวจให้คะแนนแบบพื้นฐานที่วิเคราะห์ตาม GRM และ PCM (pential credit model) กับข้อมูลจำลอง (simulated examinees) พบรการตรวจให้คะแนนแบบพื้นฐานที่วิเคราะห์ตาม GRM มีความแม่นยำในการประมาณค่าพารามิเตอร์ข้อมูลมากกว่าการวิเคราะห์ตาม PCM ซึ่งเป็นโมเดลดั้งเดิมที่มีไม่ได้พัฒนาเป็น GPCM เหมือนในการศึกษาในครั้งนี้ อย่างไรก็ดีการที่ GPCM ให้คำพ้องมากขนาดที่ต่ำกว่า GRM ก็ไม่ได้หมายความว่า GPCM เป็นโมเดลที่ไม่เหมาะสม มารุกะ (Muraki, 1993) ได้เคยศึกษาพ้องกลุ่มเส้นทางของ GPCM ในการทำประโยชน์แบบสอบของ NAEP (National Assessment of Educational Progress) และให้ข้อเสนอแนะว่าการวิเคราะห์ตาม GPCM สามารถทำให้คำพ้องมากขนาดที่ต่ำกว่า GRM โดยมีการแบ่งกลุ่ม (collapsing) ลำดับขั้นคะแนนที่มีความยากจากกลุ่เดียวกันเข้าด้วยกันหรือเรียงลำดับขั้นคะแนน (reordering categorical response) ใหม่ตามลำดับขั้นของความยากเพื่อให้มีความเหมาะสมกับ GPCM ซึ่งผลการวิเคราะห์แบบวัดคุณลักษณะตาม GPCM ในการวิจัยครั้งนี้พบว่ามีการเรียงลำดับขั้นคะแนนที่ไม่เหมาะสมต่ำกว่า GPCM (ตารางที่ 7 และภาพที่ 5-6) ดังนั้นจึงควรศึกษาเพิ่มเติมในข้อมูลอื่น ๆ โดยมีการแบ่งกลุ่มหรือเรียงลำดับขั้นคะแนนใหม่แล้ววิเคราะห์ตาม GPCM และเปรียบเทียบผลการวิเคราะห์ตาม GRM ที่ไม่ต้องปรับลำดับขั้นคะแนน ว่าจะให้ผลแตกต่างกับการศึกษาครั้งนั้นหรือไม่

จากผลการศึกษาที่ผ่านมาและการศึกษาในครั้งนี้จะเห็นได้ว่า ไม่ว่าจะเป็นการศึกษาแบบเรียงแบบวัดคุณลักษณะและแบบวัดผลลัพธ์ที่ก็ตาม การตรวจให้คะแนนแบบพื้นฐานที่วิเคราะห์ตาม GRM เป็นวิธีที่เหมาะสมกว่า และได้ข้อยุติประกอบการหนึ่งว่าหากนั้นเครื่องมือต่าง ๆ ใช้กับแบบวัดผลยานในวิเคราะห์ที่มีความสามารถสูงมาก ๆ การตรวจให้คะแนนและวิเคราะห์ตามโมเดลใดก็ตามอาจให้ผลการวิเคราะห์ที่ไม่แตกต่างกันมากนัก แต่หากใช้กับกลุ่มที่มีความสามารถต่ำมาก การใช้วิธีการตรวจให้คะแนนแบบพื้นฐานจะเป็นวิธีที่เหมาะสมกว่า
ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำไปใช้

1.1 การใช้แบบวัดคุณลักษณะกับนักเรียนที่มีระดับของคุณลักษณะปากทางต่าง (θ ตั้งแต่ -2 ถึง 0) การตรวจสอบให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM เป็นรูปที่เหมาะสมกว่า หากนำไปใช้กับนักเรียนที่มีระดับของคุณลักษณะปากทางต่าง (θ ตั้งแต่ 0 ถึง 1.5) การตรวจสอบให้คะแนนแบบรวดวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 3 อาการผิด ควรเป็นรูปที่เหมาะสม หากใช้กับนักเรียนที่มีระดับของคุณลักษณะปากทางต่าง (θ ตั้งแต่ 2 ขึ้นไป) การตรวจสอบให้คะแนนและวิเคราะห์ตามโมเดลโลจิสติก จะได้ผลที่ไม่แตกต่างกันมากนัก อย่างไรก็ตามหากต้องการนำไปใช้แบบวัดคุณลักษณะปากทางต่างที่ไม่ใช้ แต่ไม่ทราบระดับของคุณลักษณะก็ล้มด้วยอย่าง การตรวจสอบให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM จะเหมาะสมกว่า

1.2 การใช้แบบสอบถามคิดศาสตร์กับนักเรียนที่มีความสามารถระดับใดก็ตาม การตรวจสอบให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM และ GPCM จะให้ความรู้สึกสารสนเทศที่ใกล้เคียงกัน โดยเฉพาะในช่วงความสามารถของนักเรียนปากทางต่างไปทางต่าง (θ ตั้งแต่ -2 ถึง 0.5) การตรวจสอบให้คะแนนและการวิเคราะห์ต่อกล่าวเหมาะสมกว่าการตรวจสอบให้คะแนนแบบรวดวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 อาการผิด หากนำไปใช้กับนักเรียนที่มีความสามารถปากทางต่าง (θ ตั้งแต่ 0.5 ถึง 2) การตรวจสอบให้คะแนนแบบรวดวิภาคที่วิเคราะห์ตามโมเดลโลจิสติก 1, 2 และ 3 อาการผิด ควรเป็นรูปที่เหมาะสม แต่ยาไม่ทราบความสามารถของกลุ่มตัวอย่าง การตรวจสอบให้คะแนนแบบพหุวิภาคที่วิเคราะห์ตาม GRM และ GPCM จะเหมาะสมกว่า

2. ข้อเสนอแนะในการวิจัย

2.1 โปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์ข้อมูลในการวิจัยครั้งนี้มีข้อจำกัดในการวิเคราะห์ตาม GRM ด้วยโปรแกรม MULTILOG ให้คำพังก์ขั้นสารสนเทศของแบบวัดเฉพาะในช่วง θ ตั้งแต่ 2 ถึง +2 แต่ภายหลังผู้วิจัยพบว่ามีผู้พังก์ไม่ใช้งานโปรแกรม GR-GRAPH สำหรับใช้ร่วมกับ MULTILOG ซึ่งสามารถวิเคราะห์ให้คำพังก์ขั้นสารสนเทศในช่วง θ ที่กว้างขึ้น คือตั้งแต่ -5 ถึง +5 ตั้งแต่จะใช้ความเป็นมาติการศึกษาในที่อย่างเดียวกันนี้ แต่ถ้าข้อในช่วง θ ที่กว้างขึ้นกว่านั้น

2.2 การวิจัยครั้งนี้ไม่ได้กำหนดให้เครื่องมือที่ใช้เป็นตัวแปรตัวในการวิจัย ซึ่งจากผลการวิจัยพบว่าคำพังก์ขั้นสารสนเทศของแบบวัดเมื่อใช้เครื่องมือที่ต่างกันให้คำที่ต่างกันด้วย
ตั้งนี้ังครวจหรมีการศึกษาโดยทำนนคั้นให้เครื่องมือที่ใช้เป็นตัวแปรต้นในการวิจัย แล้วดูนี้ว่าในเครื่องมือที่ใช้ต่างกันจะทำให้พักย์ข้อสารสนเทศต่างกันตัวหรือไม่

2.3 แบบวัดคู่ตัวกันที่ศึกษาในครั้งนี้ใช้วิธีการจดจำให้คะแนนแบบพิษุภิก	ที่มีลักษณะเป็นมาตรประมาณค่า (1, 2, 3, 4) อยู่แล้ว และสุ่มจัดให้ราววิธีลากลำการพิริบ	เกี่ยวกับการให้คะแนนแบบทวิภิกการโดยปรับให้คะแนนในลำตัวขั้นที่ 4 เป็น 1 คะแนน และ	ลำตัวขั้นอื่น ๆ เป็น 0 คะแนน ซึ่งเป็นข้อจำกัดที่ผู้วิจัยใช้ข้อมูลที่มีผู้รวบรวมไวแล้ว ตั้งนั้น	ควรศึกษาว่าหากนั่นแบบวัดที่ใช้มาสร้างให้มีการให้คะแนนแบบทวิภิก (0-1) จริง ๆ ไม่ใช่
	เกิดจากการบูรณาการคะแนนจากมาตรประมาณค่าต่างที่ได้ปรากฏในการวิจัยครั้งนี้ จะให้ผลที่	สอดคล้องกันหรือไม่

เอกสารอ้างอิง

เอกสarerภาษาไทย

บุญหลวง บุญลาศ. 2534. การสร้างแบบสอบถามพิสัยกลุ่มสร้างเสริมกลุ่มสารสนเทศระดับ	ขั้นประกอบศึกษา. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหา-
	วิทยาลัย.

บุญรัตน์ บุญราษฏร. 2536. การศึกษาความสัมพันธ์ระหว่างความท้อถอยกับพฤติกรรมของ	ช่างเสริม สำนักงานเขตพระนครที่ 1, 2 และ 3 สำนักของสำนักงานกิจการยุทธ.
	วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิตมหาวิทยาลัยศรีนครินทรวิโรฒ สำนักมณฑ.
	วิศิ เล็กเสด็จ. 2539. การพัฒนาแนวคิดวิจัยการดำเนินการศึกษาที่เป็นจริงในการแก้ไข	ปัญหาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6. วิทยานิพนธ์ปริญญา	ครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย. อยู่ในระหว่างการจัดพิมพ.

วิชัย นิยมธรรม. 2536. ผลการใช้เทคนิคแบบจำลองทางภาคเพื่อพัฒนาเร่งด่วนธรรม	ใหม่รักษาชั่นอนุบาลปีที่ 2. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต มหา-
	วิทยาลัยศรีนครินทรวิโรฒ สำนักมณฑ.

สุทธิ ชชชวิทย์. 2534. การพัฒนาแบบวัดความข้อสัมพันธ์ของนักเรียนชั้นประถมศึกษา	ปีที่ 6. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.

สุทิดา เทียนอุดม. 2536. การพัฒนาแบบวัดความมีระเบียบวินัยสำหรับนักเรียนชั้น	ประถมศึกษาปีที่ 6. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต. จุฬาลงกรณ์มหา-
	วิทยาลัย.
ภาษาอังกฤษ


