

วนิดา ภู่อี่ยม: การเปรียบเทียบสัมประสิทธิ์การสุรอ้างอิงและความแปรปรวนของความคลาดเคลื่อนระหว่างแบบสอบวินิจฉัย
วิชาคณิตศาสตร์แบบประเพณีนิยมกับแบบสองระดับ. (A COMPARISON OF GENERALIZABILITY COEFFICIENT AND
ERROR VARIANCE BETWEEN TRADITIONAL AND TWO-TIER DIAGNOSTIC MATHEMATICS TESTS)

อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร.เอมอร จังศิริพรกรณ์, 185 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับ 2) เพื่อเปรียบเทียบสัมประสิทธิ์การสุรอ้างอิงและความแปรปรวนของความคลาดเคลื่อนระหว่างแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับที่มีจำนวนมวลเนื้อหาและจำนวนข้อสอบต่อมวลเนื้อหาที่ต่างกัน กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษา ปีที่ 1 สำนักงานเขตพื้นที่การศึกษากรุงเทพมหานคร ปีการศึกษา 2550 จำนวน 686 คน และครูผู้สอนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 จำนวน 8 คน เครื่องมือที่ใช้ในการวิจัยครั้งนี้แบ่งออกเป็น 4 ส่วน ดังนี้ ส่วนที่ 1 เครื่องมือสำรวจมโนทัศน์เรื่องการบวกเลขจำนวนเต็มลบ ส่วนที่ 2 เครื่องมือวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนเป็นแบบสอบวินิจฉัยวิชาคณิตศาสตร์ ส่วนที่ 3 เครื่องมือตรวจสอบคุณภาพด้านความตรงตามเกณฑ์สัมพันธของแบบสอบวินิจฉัยโดยใช้วิธีการคิดออกเสียง (thinking aloud) ส่วนที่ 4 เครื่องมือสำรวจความคิดเห็นในการนำแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับไปใช้ในการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อน วิเคราะห์ข้อมูลโดยการหาค่าสถิติพื้นฐาน ตรวจสอบคุณภาพด้านความตรงเชิงเนื้อหา ความตรงเชิงวินิจฉัย ความตรงตามเกณฑ์สัมพันธ ด้านความเที่ยงแบบความสอดคล้องภายใน ประมาณค่าสัมประสิทธิ์สุรอ้างอิง โดยใช้โปรแกรมสำเร็จรูป GENOVA และวิเคราะห์เนื้อหา

ผลการวิจัยพบว่า

1. แบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับที่พัฒนาขึ้น ประกอบด้วย 3 มวลเนื้อหา มวลเนื้อหาละ 4 ข้อ ตัวลวงเป็นตัวแทนมโนทัศน์การคิดวิชาคณิตศาสตร์เรื่องการบวกเลขจำนวนเต็ม
2. คุณภาพของแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับที่พัฒนาขึ้น มีคุณภาพพิจารณาตามรายข้อและรายฉบับ พบว่า

2.1 คุณภาพรายข้อของแบบสอบวินิจฉัยแบบประเพณีนิยมและแบบสอบวินิจฉัยแบบสองระดับมีค่าเฉลี่ยความยากเท่ากับ 0.63 และ 0.53 ค่าเฉลี่ยอำนาจจำแนกเท่ากับ 0.54 และ 0.51 ความเที่ยงแบบสอดคล้องภายในมีค่าเท่ากับ 0.84 และ 0.82 ตามลำดับ แสดงว่าแบบสอบวินิจฉัยแบบประเพณีนิยมง่าย สามารถจำแนกนักเรียนในระดับดี มีความคงที่ในการวัดสูง ส่วนแบบสอบวินิจฉัยแบบสองระดับค่อนข้างง่าย สามารถจำแนกนักเรียนในระดับดีมีความคงที่ในการวัดสูง

2.2 คุณภาพรายฉบับของแบบสอบวินิจฉัยแบบประเพณีนิยมและแบบสอบวินิจฉัยแบบสองระดับ พิจารณาค่าความตรงตามเกณฑ์สัมพันธของผลที่ได้จากแบบสอบวินิจฉัยวิชาคณิตศาสตร์กับผลการวินิจฉัยด้วยวิธีคิดออกเสียง พบว่า แบบสอบวินิจฉัยแบบประเพณีนิยมมีค่าเท่ากับ 0.601 ส่วนแบบสอบวินิจฉัยแบบสองระดับมีค่า 0.724 ตามลำดับ แสดงว่าแบบสอบวินิจฉัยแบบสองระดับสามารถวินิจฉัยมโนทัศน์ได้ถูกต้องแม่นยำกว่าแบบสอบวินิจฉัยแบบประเพณีนิยม พิจารณาค่าสัมประสิทธิ์อ้างอิงของแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมมีเท่ากับ 0.811 ส่วนแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบสองระดับมีค่าเท่ากับ 0.789 แสดงว่าแบบสอบวินิจฉัยแบบประเพณีนิยมมีความคงที่ในการวัดสูงกว่าแบบสอบวินิจฉัยแบบสองระดับ

3. จากการสัมภาษณ์ครู พบว่าแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมและแบบสองระดับที่พัฒนาขึ้นมีความเหมาะสมและความเป็นไปได้ในการนำไปใช้วินิจฉัยมโนทัศน์ของนักเรียน

4. เมื่อเปรียบเทียบค่าสัมประสิทธิ์การสุรอ้างอิงสอบของแบบสอบวินิจฉัยวิชาคณิตศาสตร์พบว่า แบบประเพณีนิยมมีค่าสูงกว่าแบบสองระดับทุกช่วงมวลเนื้อหา ส่วนความแปรปรวนของความคลาดเคลื่อนของแบบสอบวินิจฉัยวิชาคณิตศาสตร์แบบประเพณีนิยมมีค่าต่ำกว่าแบบสองระดับทุกช่วงมวลเนื้อหา แสดงว่าคะแนนที่สังเกตได้ของแบบสอบวินิจฉัยแบบประเพณีนิยมมีความคงที่ใกล้เคียงกับคะแนนจริงมากกว่าคะแนนที่สังเกตได้ของแบบสอบวินิจฉัยแบบสองระดับ

KEY WORD: GENERALIZABILITY COEFFICIENT / ERROR VARIANCE / TRADITIONAL DIAGNOSTIC TESTS / TWO-TIER DIAGNOSTIC TESTS / MATHEMATICS

WANIDA BHU-IAM: A COMPARISON OF GENERALIZABILITY COEFFICIENT AND ERROR VARIANCE BETWEEN TRADITIONAL AND TWO-TIER DIAGNOSTIC MATHEMATICS TESTS. THESIS PRINCIPAL ADVISOR: ASSOC. PROF. AIMORN JANGSIRIPORNPAKORN, 185 pp.

The objectives of the research were 1) to improve Mathematic Traditional Diagnostic Test and Mathematic Two-Tier Diagnostic Test 2) to compare generalizability coefficient and error variance of the Mathematic Traditional Diagnostic Test and Mathematic Two-Tier Diagnostic Test which had different test items and content mass. The samples were 686 Mathayom 1 students of Bangkok educational service area in the educational year 2007 and 8 mathematic teachers teaching in Mathayom 1. The research instruments were divided into four parts as follows: 1) sum of integer misconceptual survey 2) Mathematic Diagnostic Test for diagnosis inconsistent concept 3) the tool for checking quality of criterion-related validity of diagnostic test by using Thinking Aloud method 4) questionnaire about using Mathematic Traditional Diagnostic Test and Mathematic Two Tier Diagnostic Test to diagnose inconsistent concept. Data were analyzed by using basic statistics and checking the quality of three dimensions; criterion-related validity, estimate validity, and generalization coefficient by using GENOVA. The findings were as follows:

1. Improved Mathematic Traditional Diagnostic Test and Two-Tier Diagnostic Test consisted of three content mass. Each one was included three items. The delusion was the represent of the Mathematic Thinking Concept in sum of integer.
2. The quality of Improved Mathematic Traditional Diagnostic Test and Two-Tier Diagnostic Test were as follows:
 - 2.1 The average of difficulty of Traditional Diagnostic Test and Two-Tier Diagnostic Test were 0.63 and 0.53 respectively. The average of discrimination were 0.54 and 0.51 respectively. Internal correlation reliability were 0.84 and 0.82 respectively. The results showed that Traditional Diagnostic Test was easy, good discriminated and high stability in measurement. Two-Tier Diagnostic Test was rarely easily, good discriminated and high stability in measurement.
 - 2.2 The criterion-related validity of the Mathematic Diagnostic Test results and Thinking Aloud Method Diagnostic results was 0.601 and 0.724 respectively and indicated that Two-Tier Diagnostic Test was more precise and accurate than Traditional Diagnostic Test. The generalizability coefficient of Mathematic Traditional Diagnostic Test was 0.811 whereas Two-Tier Diagnostic Test was 0.789. The results indicated that Traditional Diagnostic Test was more stability in measurement than Two-Tier Diagnostic Test.
3. Interviewing with teachers, the results showed that Improved Mathematic Diagnostic test and Two Tier Diagnostic test were appropriate and potential to use for diagnosis students' concepts.
4. Comparing generalizability coefficient, Traditional Diagnostic test was higher than Two Tier Diagnostic Test in every levels of content mass. Moreover, the content error variance of Mathematic Traditional diagnostic test was lower than Two Tier Diagnostic test in every levels of content mass.