

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ตามระเบียบกระทรวงศึกษาธิการว่าด้วยผลการประเมินผลการศึกษาในระดับประถมศึกษา พ.ศ. 2520 ที่ใช้อยู่ในปัจจุบัน กำหนดให้มีการประเมินผลการเรียนของนักเรียนเป็นรายจุดประสงค์การเรียนรู้ซึ่งมีผลต่อการเลื่อนชั้นของนักเรียนด้วย ลักษณะการประเมินผลเป็นรายจุดประสงค์การเรียนรู้จะเป็นไปตามทฤษฎีการเรียนรู้เพื่อรอบรู้ (Mastery Learning Theory) กล่าวคือ การตัดสินใจให้นักเรียนผ่านจุดประสงค์การเรียนรู้หนึ่ง ๆ ได้ จะต้องเป็นการผ่านแบบผู้รอบรู้ในจุดประสงค์การเรียนรู้หนึ่ง ๆ ในลักษณะเช่นนี้ถ้าครูผู้สอนปฏิบัติหน้าที่ตามทฤษฎีการเรียนรู้เพื่อรอบรู้ได้เป็นอย่างดีและมีการประเมินผลโดยใช้เกณฑ์ความรอบรู้อย่างแท้จริงแล้ว ย่อมทำให้การศึกษาในระดับประถมศึกษามีคุณภาพสูง แต่ในการปฏิบัติจริงทำให้เป็นเช่นนั้นได้ยาก ดังผลการวิจัยของสถาบันวิจัยพัฒนาบริหารศาสตร์และอื่น ๆ (2526 หน้า 70-77) สามารถสรุปได้ว่า ปัจจัยที่ส่งผลต่อคุณภาพของโรงเรียนจากการพิจารณาผลสัมฤทธิ์ทางการเรียนเป็นตัวแปรตาม มี 6 ด้าน ได้แก่ ปัจจัยด้านระบบบริหารการศึกษา ปัจจัยด้านครูใหญ่ ปัจจัยด้านครู ปัจจัยด้านชุมชน ปัจจัยด้านผู้ปกครองและนักเรียน และปัจจัยด้านการเรียนการสอนและการวัดผล โดยที่ปัจจัยด้านการเรียนการสอนและการวัดผลจะส่งผลตรงที่สุดต่อผลสัมฤทธิ์ทางการเรียนของนักเรียน นอกจากนั้นจะส่งผลทางอ้อม ผู้วิจัยให้รายละเอียดด้านการเรียนการสอนและการวัดผล สรุปได้ว่า ยังขาดการเตรียมพร้อมให้แก่ครูเพื่อการสอนตามหลักสูตรครูไม่เข้าใจหลักสูตร ตั้งเป้าหมายการสอนของตนตั้งแต่เป็นเป้าหมายเฉพาะหน้า บางคนไม่มีเป้าหมายในการสอน ใช้อุปกรณ์การสอนน้อยและส่วนใหญ่ใช้วิธีสอนแบบเก่า ใช้อารมณ์ในการสอน

มากไป มอบหมายงานให้เด็กทำไม่รัดกุมเช่นสั่งแบบฝึกหัดผิดหน้าและไม่ติดตามว่าเด็กทำงานที่มอบหมายหรือเปล่า สอนโดยเน้นเฉพาะเนื้อหาของกลุ่มวิชาทักษะมีการใช้เวลาสอนมากกว่าที่กำหนด ในทุกระดับชั้น ประมาณ 20-30% ใช้เวลาเพื่อสอนไม่เต็มที่ ในด้านวัดผลพบว่าครูไม่สันทัดในการวัดผลแบบใหม่ มีการวัดผลน้อยมาก เอกสารที่ครูใช้ศึกษาเพื่อวัดผลยังไม่ละเอียดพอแก่การนำไปใช้ ครูมีงานมากจนล้นเมื่อจึงวัดผลโดยละเอียดไม่ได้ จึงมีครูบางคนจะถือว่า เด็กผ่านจุดประสงค์โดยไม่มีการวัดผลจริง ไม่ได้วัดผลให้รู้ชัดเจนว่าเด็กคนไหนมีปัญหาในวิชาอะไร อาจรู้แต่เลา ๆ ว่าเด็กคนไหนอ่อนคนไหนเก่งเท่านั้น ผลการวิจัยดังกล่าวสอดคล้องกับข้อสรุปของกรมวิชาการ (2523 หน้า 43-48) ที่สรุปผลจากการวิเคราะห์รายงานการนิเทศและติดตามผลการใช้หลักสูตรประถมศึกษา พ.ศ. 2521 จาก 51 จังหวัด พบปัญหาใกล้เคียงกันกับผลการวิจัยดังกล่าวข้างต้น

จากสภาพปัญหาดังกล่าว ถ้าจะทำให้การวัดผลและประเมินผลการศึกษามีประสิทธิภาพ จำเป็นจะต้องมีการพัฒนาเครื่องมือวัดผลทางการศึกษาที่มีคุณภาพสูง ใช้เวลาในการดำเนินการวัดผลน้อยและวิธีใช้ไม่ยุ่งยากจนเกินไปก็จะช่วยลดปัญหาการวัดผลและประเมินผลการศึกษา ดังกล่าวได้และเครื่องมือวัดผลการศึกษาที่ใช้เป็นหลักอยู่ในโรงเรียนก็คือแบบทดสอบ จึงนำที่จะพัฒนาแบบทดสอบและวิธีดำเนินการสอบเป็นสำคัญ แต่เนื่องจากการพัฒนาแบบทดสอบให้มีคุณภาพตามทฤษฎีการทดสอบดั้งเดิม (The Classical Test Theory) และวิธีดำเนินการสอบแบบเดิม ๆ จะไม่เหมาะสมสำหรับการประเมินผลเป็นรายจุดประสงค์ตามทฤษฎีการเรียนรู้ ทั้งนี้เพราะมีปัญหาสำคัญ 3 ประการดังที่ แฮมเบิลตัน (Hambleton, 1978 อ้างใน สำเร็จ บุญเรืองรัตน์ 2527 หน้า 97-98) กล่าวไว้ดังนี้

1. คุณสมบัติของข้อคำถามของแบบทดสอบไม่คงที่ แปรเปลี่ยนไปตามกลุ่มผู้สอบ ซึ่งขัดกับความต้องการที่ว่าเครื่องมือวัดผลทางวิทยาศาสตร์ใด ๆ นั้นควรมีคุณสมบัติคงที่เสมอเมื่อทดสอบกับกลุ่มใด ๆ ก็ตาม
2. ในการเปรียบเทียบความสามารถของผู้สอบต่าง ๆ นั้น จะต้องใช้แบบทดสอบที่มีข้อคำถามเหมือนกันหรือเป็นแบบทดสอบคู่ขนาน แต่การทดสอบบางครั้งนั้น เราเลือกข้อสอบมาจำนวน

หนึ่ง เพื่อวัดคุณลักษณะใดลักษณะหนึ่ง ซึ่งกรณีนี้ผู้สอบสองคนก็ไม่จำเป็นต้องสอบด้วยกลุ่มคำถามเหมือนกัน และเพราะว่าข้อสอบทั้งสองประเภทต่างก็วัดคุณสมบัติอันเดียวกัน แต่ก็อาจมีความยากง่ายไม่เท่ากัน คนหนึ่งอาจได้คำถามที่ง่าย แต่อีกคนหนึ่งได้คำถามที่ยาก ฉะนั้นคะแนนของสองคนนี้ก็ไม่สามารถนำมาจัดอันดับเพื่อเปรียบเทียบความสามารถกันได้

3. ข้อมูลที่ได้จากการทดสอบตามมาตรฐานดั้งเดิมนั้น ไม่สามารถบอกได้ว่าผู้สอบมีความสามารถทำอะไรได้ หรือมีข้อบกพร่องตอนใด เนื่องจากนำผลการสอบไปเปรียบเทียบในกลุ่มผู้สอบนั่นเอง

การแก้ปัญหาดังกล่าว จำเป็นต้องพัฒนาทฤษฎีและวิธีปฏิบัติด้านการวัดผลและประเมินผลด้วย ทฤษฎีคุณลักษณะแฝง เป็นทฤษฎีใหม่ที่ช่วยแก้ปัญหาดังกล่าว ที่เกิดขึ้นในการใช้ทฤษฎีการทดสอบดั้งเดิมได้หลายประการ ดัง โกวิท ประवालพฤษ์ (อ้างใน สำเร็จ บุญเรืองรัตน์ 2527 หน้า 130-131) กล่าวว่า "ข้อสอบที่วิเคราะห์ได้จากทฤษฎีคุณลักษณะแฝงมีคุณสมบัติ 3 ประการ ดังนี้

1. ค่าอำนาจจำแนก และค่าความยากง่ายของข้อสอบ ไม่ขึ้นอยู่กับกลุ่มตัวอย่าง
2. ข้อสอบแต่ละข้อสามารถเปรียบเทียบความยากง่ายกัน โดยอิสระ
3. ค่าสถิติที่อธิบายคุณภาพของข้อสอบ จากผลการวิเคราะห์ด้วยทฤษฎีคุณลักษณะแฝง นั้นนับว่าเพียงพอแล้ว ไม่ต้องอาศัยค่าอื่น ๆ เพิ่มเติม

จากคุณสมบัติทั้งสามประการนี้จึงสามารถสร้างข้อสอบที่รวมกลุ่มคำถามได้อย่างประหยัดและมีประสิทธิภาพ" ดังนั้นการพัฒนารูปแบบทดสอบจึงสมควรใช้วิธีวิเคราะห์ข้อสอบตามทฤษฎีคุณลักษณะแฝงดังกล่าว จะได้เป็นการแก้ปัญหาดังกล่าวที่มีอยู่ให้ลดลงไป

อีกปัญหาหนึ่งที่ไม่สามารถกระทำได้ โดยทฤษฎีการทดสอบดั้งเดิมก็คือปัญหาการจัดข้อสอบให้สอดคล้องกับความสามารถของบุคคล กล่าวคือ การทดสอบดั้งเดิมนั้นจะใช้ได้อย่างเหมาะสมในกรณีที่ผู้สอบมีความสามารถในระดับกลางหรือในช่วงใกล้ค่าเฉลี่ยของกลุ่มผู้สอบ ดังนั้นผู้สอบที่มีความสามารถสูงหรือต่ำกว่าระดับเฉลี่ยของกลุ่มมาก ๆ จะให้รายละเอียดเกี่ยวกับตัวผู้สอบต่ำ (Vale and Weiss, 1975 : 6) ในกรณีเช่นนี้ถ้าผู้สอบได้ตอบข้อสอบที่มีระดับความยากสูงกว่าระดับความสามารถของตนแล้ว ผู้สอบมีแนวโน้มจะตอบข้อสอบโดยการเดาสูงขึ้นและถ้าผู้สอบคนใดได้ตอบข้อสอบที่มีระดับความยากต่ำกว่าระดับความสามารถของตนเองแล้ว ข้อสอบนั้นจะ ไม่ช่วยๆ ให้ผู้สอบได้ใช้ความสามารถในระดับสูงสุด ซึ่งมีผลให้ความเชื่อมั่นของคะแนนผลการสอบจากผู้สอบทั้งสองกลุ่มต่ำกว่าที่ควรจะเป็น (Weiss, 1974 : 1) ปัญหาดังกล่าวสามารถแก้ไขได้อย่างเหมาะสมโดยเทคนิคการจัดการทดสอบแบบเทเลอร์ (Tailored Testing) ซึ่งเป็นตัวอย่างหนึ่งของการประยุกต์ใช้ทฤษฎีคุณลักษณะแฝงอย่างได้ผล (Rudner, 1983 : 951) ในกระบวนการของการทดสอบแบบเทเลอร์ จะเลือกข้อสอบที่มีระดับความยากเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน การสอบครั้งหนึ่ง ๆ ผู้สอบแต่ละคนไม่จำเป็นต้องตอบข้อสอบเหมือนกันทุกข้อทั้งฉบับ แต่ขึ้นอยู่กับผลการตอบข้อสอบในแบบทดสอบของแต่ละคน (Weiss and Kingsbury, 1984 : 361) โดยทั่ว ๆ ไปแล้วจะเริ่มจากข้อสอบที่มีความยากระดับปานกลาง ถ้าผู้สอบตอบข้อสอบข้อแรกถูกต้อง ข้อสอบข้อต่อไปก็จะยากขึ้น แต่ถ้าตอบข้อสอบผิดข้อต่อไปก็จะง่ายลง การทำข้อสอบข้อต่อไปจะปฏิบัติเช่นเดียวกันนั้นสิ้นสุดการสอบ. (Green and others, 1984 : 347)

จากการศึกษาของเบทซ์ และไวส์ (Betz and Weiss, 1973 Cited by Larkin and Weiss, 1975 : 1) พบว่าการทดสอบแบบเทเลอร์จะใช้ข้อสอบน้อยข้อกว่าการทดสอบแบบดั้งเดิม จึงใช้เวลาในการทำแบบทดสอบน้อยลง ในขณะที่ความเชื่อมั่นและความเที่ยงตรงคงเดิม

จะเห็นได้ว่าการพัฒนาข้อสอบตามทฤษฎีคุณลักษณะแฝง จะช่วยแก้ปัญหาในด้านคุณภาพข้อสอบ ในขณะที่การทดสอบแบบเทเลออร์ช่วยเพิ่มคุณภาพของการทดสอบและช่วยลดเวลาในการดำเนินการสอบด้วย จึงน่าจะเป็นกระบวนการที่ช่วยให้การวัดผลและประเมินผลมีประสิทธิภาพดีขึ้น ในปัจจุบันนี้ การทดสอบแบบเทเลออร์ที่ใช้อย่างได้ผลในประเทศสหรัฐอเมริกาจะใช้คอมพิวเตอร์เป็นเครื่องมือในการประเมินค่าความสามารถของนักเรียน และ เลือกข้อสอบที่เหมาะสมออกมาให้นักเรียนตอบในข้อต่อไป ซึ่งคงไม่สามารถดำเนินการโดยคอมพิวเตอร์ได้ในประเทศไทยในปัจจุบัน แต่อย่างไรก็ตาม มีการจัดแบบทดสอบแบบเทเลออร์หลายรูปแบบที่สามารถจัดเป็นแบบกระดาษดินสอ (Paper - pencil test) ได้ ดังนั้น ผู้วิจัยจึงสนใจที่จะนำรูปแบบดังกล่าวมาบางรูปแบบมาทดลองใช้ในการตัดสินความรอบรู้ตามจุดประสงค์การเรียนรู้ ถ้าผลการตัดสินมีความเที่ยงตรงสูงเมื่อเทียบกับการทดสอบแบบเทเลออร์ตามกระบวนการที่ใช้ในคอมพิวเตอร์ ก็จะได้รูปแบบของแบบทดสอบที่จะดำเนินการได้ในประเทศไทยและช่วยให้การวัดผลและประเมินผลการศึกษาในโรงเรียนประถมศึกษาที่มีประสิทธิภาพมากยิ่งขึ้น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อศึกษาความเที่ยงตรงของการตัดสินความรอบรู้ตามจุดประสงค์การเรียนรู้ในวิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ตามผลการทดสอบด้วยแบบทดสอบแบบเทเลออร์ 3 รูปแบบคือ แบบปิรามิดที่ใช้ขนาดขั้นคงที่ (Constant Step Size Pyramid Model) จำนวน 6 ขั้นและ 10 ขั้น แบบปิรามิดที่มีข้อสอบหลายข้อในแต่ละขั้น (Multiple Item Pyramid Model) จำนวน 3 ขั้นและ 5 ขั้น และแบบเปลี่ยนระดับ (Flexilevel Model) จำนวน 6 ขั้นและ 10 ขั้น
2. เพื่อเปรียบเทียบความแตกต่างของผลการตัดสินความรอบรู้ระหว่างแบบทดสอบแบบเทเลออร์ 3 รูปแบบ ที่มีจำนวนข้อที่นักเรียนต้องตอบเท่ากัน กับผลการตัดสินตามปกติของครู

3. เพื่อเปรียบเทียบความแตกต่างของผลการตัดลึนความรอบรู้ระหว่างแบบทดสอบแบบ
 เทเลอร์ รูปแบบเดีวกันแต่มีจำนวนข้อที่นักเรียนต้องตอบต่างกัน

ประโยชน์ของการวิจัย

ผลจากการวิจัยจะก่อให้เกิดประโยชน์ดังต่อไปนี้

1. ช่วยให้ทราบว่าแบบทดสอบแบบเทเลอร์ 3 รูปแบบ เมื่อนำมาประยุกต์ใช้กับการ
 ตัดลึนความรอบรู้ตามจุดประสงค์การเรียนรู้ในวิชาคณิตศาสตร์ จะมีความเที่ยงตรงในการตัดลึน
 หรือไม่
2. ช่วยให้ทราบว่าผลการตัดลึนโดยใช้ผลการสอบจากแบบทดสอบแบบเทเลอร์ 3 รูป
 แบบกับผลการตัดลึนตามปรกติของครูมีความแตกต่างกันหรือไม่
3. ช่วยให้ทราบว่า รูปแบบการทดสอบแบบเทเลอร์ที่ทดลองใช้ครั้งนี้ รูปแบบใดให้ผล
 การตัดลึน ใกล้เคียงกับผลการตัดลึนความรอบรู้ที่ใช้เป็นเกณฑ์มากกว่ากัน เมื่อแต่ละรูปแบบมี
 จำนวนข้อที่นักเรียนต้องตอบ เท่ากัน
4. ช่วยให้ทราบว่า จำนวนข้อสอบที่นักเรียนจะต้องตอบ ในแบบทดสอบแบบเทเลอร์
 3 รูปแบบ มีผลทำให้การตัดลึนความรอบรู้เปลี่ยนแปลงไปหรือไม่อย่างไร
5. ใช้เป็นเครื่องช่วยในการตัดลึนใจว่า ควรส่งเสริมให้มีการพัฒนาแบบทดสอบแบบ
 เทเลอร์ เพื่อการประเมินผลสัมฤทธิ์ทางการเรียนหรือไม่
6. ใช้เป็นแนวทางสำหรับการศึกษาวิจัยเกี่ยวกับการทดสอบแบบเทเลอร์ในลักษณะอื่น ๆ
 ต่อไป

สมมติฐานของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ตั้งสมมติฐานไว้ดังนี้

1. ผลการตัดสินความรอบรู้ของนักเรียน เมื่อใช้ผลการสอบจากแบบทดสอบแบบเทเลอร์ ที่จัดตามแบบจำลองรูปปริมาตรที่ใช้ขนาดขั้นคงที่จำนวน 6 ขั้นและ 10 ขั้น แบบจำลองรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 3 ขั้นและ 5 ขั้น และแบบจำลองเปลี่ยนระดับจำนวน 6 ขั้นและ 10 ขั้น มีความเที่ยงตรงในการตัดสินอย่างมีนัยสำคัญทางสถิติ

2. ผลการตัดสินความรอบรู้ของนักเรียน เมื่อใช้ผลการสอบจากแบบทดสอบแบบเทเลอร์ ที่จัดข้อสอบตามแบบจำลองรูปปริมาตรที่ใช้ขนาดขั้นคงที่จำนวน 10 ขั้น แบบจำลองรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 5 ขั้น แบบจำลองเปลี่ยนระดับจำนวน 10 ขั้น และตามการทดสอบความรอบรู้แบบเทเลอร์ (AMT) จะมีผลแตกต่างจากการตัดสินตามปกติของครูในชั้นเรียน

3. ผลการตัดสินความรอบรู้ของนักเรียน เมื่อใช้ผลการสอบจากแบบทดสอบแบบเทเลอร์ ที่จัดข้อสอบตามแบบจำลองรูปปริมาตรที่ใช้ขนาดขั้นคงที่จำนวน 6 ขั้น แบบจำลองรูปปริมาตรที่มีข้อสอบหลายข้อในแต่ละขั้นจำนวน 3 ขั้น และแบบจำลองเปลี่ยนระดับจำนวน 6 ขั้น จะมีผลแตกต่างจากการตัดสินตามการทดสอบความรอบรู้แบบเทเลอร์ (AMT) และแตกต่างจากการตัดสินตามปกติของครูในชั้นเรียน

4. ผลการตัดสินความรอบรู้ของนักเรียน เมื่อใช้ผลการสอบจากแบบทดสอบแบบเทเลอร์ แบบเดียวกัน แต่มีจำนวนข้อที่นักเรียนต้องตอบต่างกัน จะมีความแตกต่างกัน

ข้อตกลงเบื้องต้นของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยยึดข้อตกลงเบื้องต้นดังต่อไปนี้

1. แบบทดสอบที่ใช้ในการวิจัยครั้งนี้ มีคุณสมบัติตามทฤษฎีคุณลักษณะแฝง 3 ประการ คือ ความเป็นมิติเดียว (Unidimensional) ข้อสอบแต่ละข้อเป็นอิสระจากกัน (Local Independence) และโอกาสที่ผู้เข้าสอบจะตอบข้อสอบถูกต้องขึ้นอยู่กับได้งลักษณะข้อสอบ ไม่ขึ้นอยู่กับ การแจกแจงความสามารถของกลุ่ม

2. ประชากรในการวิจัยมีคุณสมบัติความเป็นอิสระ (Population free)
3. ค่าความชุก (b) ค่าอำนาจจำแนก (a) ของข้อสอบ และค่าความสามารถ (e) ที่ได้จากการวิเคราะห์ข้อสอบแบบโลจิสติกที่มีพารามิเตอร์ 3 ตัว เป็นค่าการวัดในระดับอันตรภาคและมีการแจกแจงแบบปรกติ
4. ก่อนการทดสอบความรอบรู้ ถือว่านักเรียนทุกคนมีความสามารถเท่ากับความสามารถที่เป็นเกณฑ์ความสามารถขั้นต่ำของผู้รอบรู้ที่ได้จากความเห็นของคณะครูผู้สอนคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 และค่าความแปรปรวนในการประมาณค่าความสามารถก่อนเริ่มการตอบข้อสอบข้อแรก มีค่า = 1 ซึ่งเท่ากับค่าความแปรปรวนของคะแนนมาตรฐานที่

ขอบเขตของการวิจัย

1. เนื่องจากข้อตกลงด้านความเป็นมติเดียวของการวิเคราะห์ข้อสอบตามแบบจำลองโลจิสติก ผู้วิจัยได้พิจารณาคัดเลือกจุดประสงค์การเรียนรู้ในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 มาใช้ในการวิจัยครั้งนี้ เพียง 1 จุดประสงค์การเรียนรู้ โดยพิจารณาว่าเป็นจุดประสงค์การเรียนรู้ที่จะได้รับอิทธิพลจากความสามารถด้านภาษาให้น้อยที่สุด เพื่อให้เกิดคุณสมบัติความเป็นมติเดียว จุดประสงค์การเรียนรู้ที่เลือกได้คือจุดประสงค์การเรียนรู้ที่ 4 "เขียนเลขในรูปยกกำลัง ห.ร.ม. และ ค.ร.น. ได้" และเลขจำนวนที่ใช้ในแบบทดสอบจะต้องเป็นเลขจำนวนเต็มไม่เกิน 3 หลัก

2. ประชากรในการวิจัยนี้มี 2 กลุ่ม คือ

2.1 ประชากรในการพัฒนาเครื่องมือ ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2529 ของจังหวัดอุดรดิตถ์ แพร่ และน่าน

2.2 ประชากรในการศึกษาตามวัตถุประสงค์ของการวิจัย ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2531 ของโรงเรียนอนุบาลอุดรดิตถ์

3. การทดสอบแบบเทเลอร์ในการวิจัยนี้ จำกัดขอบเขตของรูปแบบการทดสอบไว้เพียง 3 รูปแบบ ๗ ละ 2 ขนาดความยาว คือ รูปแบบปิรามิดที่ใช้ขนาดชั้นคงที่จำนวน 6 ชั้นและ 10 ชั้น รูปแบบปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้นจำนวน 3 ชั้นและ 5 ชั้น และ รูปแบบเปลี่ยนระดับจำนวน 6 ชั้นและ 10 ชั้น สำหรับรูปแบบปิรามิดที่ใช้ขนาดชั้นคงที่กับรูปแบบเปลี่ยนระดับ นักเรียนต้องตอบข้อสอบชั้นละ 1 ข้อ ส่วนรูปแบบปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้น นักเรียนต้องตอบข้อสอบชั้นละ 2 ข้อ ดังนั้นจึงสรุปได้ว่า ในแต่ละรูปแบบของการทดสอบแบบเทเลอร์ที่ศึกษา มีจำนวนข้อสอบที่นักเรียนต้องตอบ 2 ขนาดความยาวคือ 6 ข้อและ 10 ข้อ

นอกจากการทดสอบแบบเทเลอร์ 3 รูปแบบดังกล่าว ผู้วิจัยได้ใช้การทดสอบแบบเทเลอร์ตามยุทธวิธีของเบย์เป็นเกณฑ์ในการตรวจสอบความเที่ยงตรงของการตัดสินของแบบทดสอบที่ศึกษา และนำผลการตัดสินตามปรกติของครูผู้สอนมาเปรียบเทียบกับผลจากการทดสอบแบบเทเลอร์ด้วย

นิยามศัพท์เฉพาะ

1. แบบทดสอบแบบเทเลอร์ หมายถึง แบบทดสอบที่จัดตามแบบจำลองรูปปิรามิดที่ใช้ขนาดชั้นคงที่ แบบจำลองปิรามิดที่มีจำนวนข้อสอบหลายข้อในแต่ละชั้นและแบบจำลองเปลี่ยนระดับ
2. ประสิทธิภาพของแบบทดสอบแบบเทเลอร์ หมายถึง การนำผลจากการทดสอบด้วยแบบทดสอบแบบเทเลอร์ไปใช้ในการตัดสินความรอบรู้ได้อย่างเที่ยงตรง โดยที่นักเรียนต้องตอบข้อสอบจำนวนน้อยข้อกว่าปกติที่สอบกันในโรงเรียน
3. แบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นคงที่ หมายถึง แบบทดสอบที่จัดตามแบบจำลองรูปปิรามิดที่มีจำนวนข้อในแต่ละชั้นเท่ากับลำดับที่ของชั้น ในแต่ละชั้นนักเรียนต้องตอบข้อสอบ 1 ข้อ และใช้กฎการแยกทางแบบเพิ่ม $1 / ลด 1$ ในการดำเนินการสอบ แบ่งเป็น 2 ฉบับคือ ฉบับที่ 1 มีจำนวนชั้น 6 ชั้น และฉบับที่ 2 มีจำนวน 10 ชั้น
4. แบบทดสอบรูปปิรามิดที่มีจำนวนข้อสอบหลายข้อในแต่ละชั้น หมายถึง แบบทดสอบที่จัดตามแบบจำลองรูปปิรามิด โดยให้แต่ละชั้นมีข้อสอบที่นักเรียนต้องตอบจำนวน 2 ข้อ และใช้กฎการแยกทางแบบตอบถูกทั้ง 2 ข้อ จะเพิ่มความยากของข้อสอบในชั้นถัดไป แต่ถ้าตอบผิดทั้ง 2 ข้อหรือตอบถูก 1 ข้อ จะลดความยากของข้อสอบในชั้นถัดไป แบ่งเป็น 2 ฉบับคือ ฉบับที่ 1 มีจำนวนชั้น 3 ชั้น และฉบับที่ 2 มีจำนวน 5 ชั้น

5. แบบทดสอบเปลี่ยนระดับ หมายถึง แบบทดสอบที่จัดตามแบบจำลองของการทดสอบเปลี่ยนระดับ โดยที่ทุกข้อในแบบทดสอบจะต้องมีค่าความยากต่างกัน ในแต่ละชั้นนัก เรียนจะต้องตอบ 1 ข้อ ถ้าหากเรียนตอบถูกจะต้องตอบข้อที่ยากขึ้นในชั้นถัดไป แต่ถ้าตอบผิดจะต้องตอบข้อที่ง่ายลงในชั้นถัดไป แบ่งเป็น 2 ฉบับคือ ฉบับที่ 1 มีจำนวนชั้น 6 ชั้น และฉบับที่ 2 มีจำนวนชั้น 10 ชั้น

6. เกณฑ์ความสามารถขั้นต่ำของผู้รอบรู้ (θ_m) หมายถึง ค่าความสามารถเฉลี่ยที่ได้จากความเห็นของครูผู้สอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่ให้พิจารณาข้อสอบ 30 ข้อ ที่สุ่มจากกลุ่มข้อสอบว่า อย่างน้อยนักเรียนที่จัดว่าเป็นผู้รอบรู้แล้วควรตอบข้อใด ได้ถูกต้องบ้าง นำความเห็นของครูแต่ละคน ไปประมาณค่าความสามารถ (θ) ตามวิธีของเบย์ท์และไวส์ส์ (Larkin and Weiss, 1975 : 10) ซึ่งตัดแปลงจากวิธีความเป็นไปได้สูงสุด (Maximum likelihood) จากนั้น จึงนำค่าความสามารถนั้นมาหาค่าเฉลี่ย ในที่มีค่าเท่ากับ 0.6329

7. การทดสอบความรู้แบบเทเลอร์ (AMT) หมายถึง กระบวนการทดสอบที่จัดให้นักเรียนได้ตอบข้อสอบที่เหมาะสมกับความสามารถของนักเรียนให้มากที่สุด ตามยุทธวิธีของเบย์ท์ โดยเริ่มสอบข้อแรกด้วยข้อสอบที่คัดเลือกจากกลุ่มข้อสอบที่มีค่าความยากใกล้เคียงที่สุดกับค่า θ_m คำตอบของนักเรียนจะจำแนกเป็นถูกหรือผิด แล้วประมาณค่าความสามารถใหม่โดยวิธีของเบย์ท์ ซึ่งเสนอโดยโอเวน (Roger J. Owen, 1975 : 353) จากนั้นจึงจัดข้อสอบข้อต่อไปให้ตอบโดยเลือกข้อที่มีค่าความยากใกล้เคียงกับระดับความสามารถใหม่ ดำเนินการเช่นนี้ไปจนกระทั่งเกิดกรณีใดกรณีหนึ่งดังต่อไปนี้ จึงยุติการสอบ

ก. ค่า θ_m หลุดพ้นจากช่วงค่าความสามารถที่ประมาณได้อย่างมีนัยสำคัญทางสถิติที่ระดับความมีนัยสำคัญ 0.5

ข. ค่าความคลาดเคลื่อนมาตรฐานในการประเมินค่าความสามารถ ที่ได้จากข้อสอบที่เลือกมาให้ให้นักเรียนตอบข้อสุดท้ายมีค่าลดลงจากข้อก่อนหน้าไม่ถึง 0.005

8. ความเที่ยงตรงในการตัดสินความรู้รอบรู้ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์แบบพาย (Phi-Correlation Coefficient : ϕ) ระหว่างผลการตัดสินจากการทดสอบความรู้รอบรู้แบบเทเลอร์ (AMT) กับผลการตัดสินจากผลการสอบด้วยแบบทดสอบตามแบบจำลองต่าง ๆ เมื่อใช้เกณฑ์ความสามารถขั้นต่ำของผู้รอบรู้ (θ_m) เป็นเกณฑ์ในการตัดสิน