

ความเป็นมาและความสำคัญของปัญหา

หลังจากได้มีการเปลี่ยนแปลงหลักสูตรตั้งแต่พุทธศักราช 2521 ระบบของการวัดและประเมินผลจึงจำเป็นต้องปรับปรุงเปลี่ยนแปลงตามไปด้วยทั้งนี้ก็เพื่อให้เกิดความสอดคล้องระหว่างการจัดการเรียนการสอนกับการประเมินผล กล่าวคือ วิธีการวัดผลเปลี่ยนแปลงจากการสอบเพื่อตัดสินได้ตก มาเป็นการวัดเพื่อพัฒนาบุคคล ซึ่งต้องการพัฒนาผู้เรียนทุกคนให้บรรลุวัตถุประสงค์ที่กำหนดไว้ หรือผู้ว่านักเรียนได้เปลี่ยนแปลงพฤติกรรมไปตามเกณฑ์ที่กำหนดไว้หรือไม่ การวัดและประเมินผลในลักษณะนี้เป็นผลสืบเนื่องมาจากการประยุกต์เอาวิธีการการเรียนแบบผ่านเกณฑ์หรือแบบรอบรู้ (Mastery Learning) มาใช้ (สุธรรม จันทน์หอม 2527 หน้า 9)

ในระดับประถมศึกษาครูผู้สอนจะต้องประเมินผลการเรียนตามจุดประสงค์การเรียนรู้ในสมุดประจำชั้น (ป.02) กล่าวคือ ต้องจำแนกนักเรียนออกได้ว่า ใครมีความรอบรู้ในเรื่องที่สอนแล้วหรือยัง ครูผู้สอนจึงจำเป็นต้องสร้างเครื่องมือมาวัดในจุดประสงค์นั้น ๆ โดยออกข้อสอบให้สอดคล้องกับจุดประสงค์ในการเรียนการสอน และใช้สอบเพื่อทราบจุดเด่น จุดด้อยของนักเรียน และมีการสอนซ่อมเสริมนักเรียนที่ไม่รอบรู้ หรือไม่บรรลุวัตถุประสงค์เชิงพฤติกรรมที่กำหนดไว้ การใช้แบบทดสอบอิงเกณฑ์ส่วนใหญ่ มักจะอ้างอิงจุดประสงค์เชิงพฤติกรรมและการรอบรู้ควบคู่กันไปหรืออาจจะกล่าวได้ว่า การทดสอบแบบอิงเกณฑ์จะนำมาใช้ในระบบการเรียนการสอนที่ยึดเอาวัตถุประสงค์เชิงพฤติกรรมเป็นหลัก และการเรียนรู้ที่ต้องการคือให้ทุก ๆ คนมีความสามารถอย่างน้อยที่สุดไม่ต่ำกว่าเกณฑ์ที่ระบุไว้ เช่นต้องเขียนแปลศัพท์ได้ถูกต้องไม่ต่ำกว่า 80 เปอร์เซ็นต์ของคำศัพท์ทั้งหมดที่กำหนดให้ (โกวิท ประวาลพฤษ์ 2523 หน้า 16-17) ผลจากการสอบก็จะทำการจำแนกหรือตัดสิน โดยให้นักเรียนที่ได้คะแนนสูงกว่าเกณฑ์ที่กำหนดเป็นผู้ที่สอบผ่านจุดประสงค์และนักเรียนที่สอบได้ต่ำกว่าเกณฑ์ก็จะเป็นผู้ไม่ผ่านจุดประสงค์

จะเห็นว่าในระบบการเรียนการสอน ครูจะต้องจำแนกหรือตัดสินผลการเรียนอยู่ตลอดเวลา ซึ่งในเรื่องนี้มักจะเกิดปัญหาอยู่เสมอว่า ครูจำแนกหรือตัดสินไม่ถูก ไม่ตรงกับความจริง เกิดความคลาดเคลื่อนขึ้น กล่าวคือ ครูจำแนกนักเรียนว่ามีความรอบรู้ทั้ง ๆ ที่ความจริงนักเรียนไม่มีความรอบรู้หรือความสามารถเพียงพอ และในบางครั้งครูก็จำแนกว่านักเรียนไม่มีความรอบรู้หรือความสามารถเพียงพอทั้ง ๆ ที่ความจริงนักเรียนมีความรอบรู้แล้ว ทั้งสองกรณีเป็นความคลาดเคลื่อนของการตัดสิน สาเหตุดังกล่าวนี้อาจเกิดจากการขาดคุณภาพด้านต่าง ๆ เช่น เครื่องมือวัดหรือแบบสอบ การดำเนินการสอบ การตรวจให้คะแนน เกณฑ์การตัดสิน (Emrick, 1977)

ดังนั้น เครื่องมือที่ใช้วัดความสามารถของนักเรียนจึงเป็นองค์ประกอบที่สำคัญอย่างหนึ่งต่อความเที่ยงตรงในการจำแนกความรอบรู้ของนักเรียน นั่นคือ หากเครื่องมือที่ใช้วัดสามารถให้ผลการจำแนกผู้สอบออกเป็นผู้รอบรู้และไม่รอบรู้ได้อย่างถูกต้องตามสภาพที่แท้จริงแล้ว ก็จะไม่ก่อให้เกิดความคลาดเคลื่อนในการตัดสินจำแนกความรอบรู้ ทั้งนี้เป็นเพราะเรายังไม่อาจสร้างมาตรวัดที่สมบูรณ์แบบได้ ซึ่งปัญหาในเรื่องนี้ก็ยังเป็นปัญหาหลักของการวัดผลในปัจจุบันเช่นเดียวกับที่บินเน็ต (Binet) และเทอร์สโตน (Thurstone) เผชิญมาแล้ว จนกระทั่งในปัจจุบันก็ต้องเผชิญปัญหาเดียวกันอีกคือเรื่องของมาตราที่ใช้วัดคุณลักษณะภายในตัวของคนนั่นเอง (สุธรรม จันทน์หอม 2528 หน้า 1)

เครื่องมือที่นำมาวัดความรู้ความสามารถที่นิยมใช้กันในปัจจุบันคือแบบทดสอบ โดยเฉพาะอย่างยิ่งแบบทดสอบเลือกตอบ ซึ่งปกติแล้วการวัดความรู้ความสามารถในเรื่องใดเรื่องหนึ่ง ผู้สอบทุกคนจะต้องสอบแบบทดสอบที่เหมือนกันและตอบข้อสอบทุกข้อที่มีอยู่ในแบบทดสอบ จะเห็นว่าการทดสอบที่ใช้เครื่องมือในลักษณะนี้มีได้ค่านิ่งว่าแบบทดสอบนั้นจะเหมาะสมกับความสามารถของผู้สอบแต่ละคนหรือไม่ ถ้าบางส่วนของแบบทดสอบยากมากเกินไปสำหรับผู้สอบคนหนึ่งแล้ว เขาก็จะรู้สึกท้อแท้และอาจไม่ทำข้อสอบในส่วนนั้น ดังนั้นเมื่อแบบทดสอบฉบับหนึ่งยากมากเกินไปกับบางส่วนของกลุ่มผู้สอบแล้วก็มีแนวโน้มว่าเขาจะทำข้อสอบโดยการเดาสูง ตรงกันข้ามถ้ามีข้อสอบบางส่วนของแบบทดสอบง่ายเกินไปกับผู้สอบอีกส่วนหนึ่ง ลักษณะเช่นนี้ข้อสอบก็จะไม่ท้าทายให้ผู้สอบได้

แสดงความสามารถสูงสุดออกมา กลับกลายเป็นสร้างความเบื่อหน่ายให้กับผู้สอบไป ทั้งสองกรณีที่กำลังมานี้คือข้อสอบที่ง่ายและยากมากเกินไปกับบางส่วนของกลุ่มผู้สอบ เป็นองค์ประกอบเบื้องต้นที่ทำให้คะแนนที่ได้จากการทดสอบมีความเที่ยงตรงต่ำลง ไม่ตรงกับความสามารถที่แท้จริงของผู้สอบ นำไปสู่การจำแนกหรือตัดสินการรอบรู้ที่คลาดเคลื่อน (Weiss, 1974 : 1)

นอกจากนี้ การกำหนดจำนวนข้อทดสอบที่ตายตัวเพื่อทดสอบกับกลุ่มผู้สอบที่มีระดับความสามารถแตกต่างกัน มีผลต่อการวัดขีดความสามารถที่แท้จริงหรือความสามารถที่ซ่อนเร้นอยู่ภายในตัวผู้สอบ ซึ่งทฤษฎีการวัดทางจิตวิทยาชี้ให้เห็นว่ามีผลกระทบต่อความเชื่อมั่นและความเที่ยงตรงของคะแนนที่ได้ ลอร์ด (Lord, 1977 : 125) ชี้ให้เห็นว่าถ้ากลุ่มผู้สอบมีความแตกต่างกันมากมันเป็นไปได้ที่แบบทดสอบดั้งเดิม (Conventional Test) ที่เราใช้กันอยู่จะวัดได้เที่ยงตรงทุก ๆ กลุ่มย่อยที่มีความสามารถสูงและต่ำในเวลาเดียวกัน จากผลการวิจัยคะแนนจากการทดสอบจะมีความเที่ยงตรงมากที่สุด เมื่อข้อสอบมีความยากง่ายพอดี ๆ กับความสามารถของผู้สอบ ซึ่งทำให้โอกาสในการตอบถูกหรือผิดนั้น 50 ต่อ 50 เปอร์เซ็นต์ (ความน่าจะเป็นในการตอบข้อสอบถูก .50) ซึ่งฮิค (Hick, 1951 cited in Weiss, 1974: 1) ก็ได้ข้อสรุปในลักษณะเดียวกัน

จากการค้นพบนี้ แสดงว่าแบบทดสอบฉบับหนึ่งจะให้ผลการวัดที่มีความเที่ยงตรงต่ำลง (คะแนนที่ได้จากการทดสอบมีความคลาดเคลื่อนมาตรฐานของการวัดสูงขึ้น) เมื่อผู้สอบคนหนึ่งมีโอกาสที่จะตอบข้อสอบข้อหนึ่ง ๆ ถูกได้น้อยมากหรือผิดอีกอย่างหนึ่งว่าเกือบไม่มีเปอร์เซ็นต์ในการตอบถูกเลย (ข้อสอบยากมาก ๆ ทำให้ความน่าจะเป็นในการตอบถูกอยู่ระหว่าง .00 ถึง .50) และข้อสอบที่ผู้สอบคนนั้นจะตอบถูกได้อย่างแน่นอน 100 เปอร์เซ็นต์ (ข้อสอบง่ายมาก ๆ ทำให้ความน่าจะเป็นในการตอบถูกมากกว่า .50 หรือใกล้ 1.00 มาก) ดังนั้นจึงปรากฏชัดว่าถ้าผู้สอบคนหนึ่งสอบแบบทดสอบที่ประกอบด้วยข้อสอบที่ง่ายมาก ๆ หรือยากมาก ๆ เกินความสามารถของเขาแล้วผลการวัดจะไม่เที่ยงตรงเท่ากับใช้แบบทดสอบที่ประกอบด้วยข้อสอบที่มีความยาก พอเหมาะพอดี (ปานกลาง) กับความสามารถของผู้สอบคนนั้น (Weiss, 1974 : 1) ซึ่งลอร์ด (Lord, 1970, 1971 cited in Weiss, 1974 : 1) สรุปถึงการใช้แบบทดสอบดั้งเดิม

คือคะแนนที่ได้จากกลุ่มผู้สอบซึ่งมีความสามารถต่ำกว่าหรือสูงกว่าความสามารถของกลุ่มนั้น จะมี ความเที่ยงตรงต่ำกว่ากลุ่มผู้สอบที่มีความสามารถกลาง ๆ ของกลุ่ม สอดคล้องกับเวลและไวส์ (Vale and Weiss 1975: 6) ที่กล่าวว่าแบบทดสอบดั้งเดิมจะใช้ได้อย่างเหมาะสมในกรณีที่ผู้ สอบมีความสามารถในระดับปานกลางหรือในช่วงใกล้ค่าเฉลี่ยของกลุ่ม ดังนั้นผู้สอบที่มีความ สามารถสูงหรือต่ำกว่าค่าเฉลี่ยของกลุ่มมาก ๆ ก็จะทำให้สาระประโยชน์ (Information) เกี่ยว กับผู้สอบต่ำ

จากเหตุผลที่กล่าวมาแล้วข้างต้นจึงมีการหาวิธีการทดสอบแบบใหม่มาใช้ เพื่อแก้ไข การทดสอบแบบเดิมให้ดีขึ้น ในอันที่จะทำให้ได้คะแนนที่มีความเชื่อมั่นและมีความเที่ยงตรงกับ คุณลักษณะความสามารถที่แท้จริงของผู้สอบ วิธีการทดสอบที่นำมาใช้คือวิธีการทดสอบเทเลอร์ (Tailored Testing) ซึ่งมีคำที่เรียกในความหมายเดียวกับการทดสอบแบบเทเลอร์ (Lord, 1980) เช่น การทดสอบแบบอะแดปทีฟ (Adaptive Testing) การทดสอบแบบแตกแขนง (Branched Testing) การทดสอบรายบุคคล (Individualized Testing) การทดสอบ ที่ให้ข้อสอบต่อเนื่องกัน (Sequential Testing) การทดสอบบนเงื่อนไขผลการตอบ (Response Contingent Testing) หรือถ้าใช้กับคอมพิวเตอร์ก็เรียกว่าการทดสอบโดย โปรแกรม (Programmed Testing) การทดสอบโดยใช้คอมพิวเตอร์ (Computerized Testing) ซึ่ง สุธรรม จันทน์หอม (2529 หน้า 10) เรียกการทดสอบเทเลอร์ว่าการ ทดสอบต่อเนื่อง

การทดสอบเทเลอร์ยึดความคิดพื้นฐานที่ว่า ข้อสอบที่จะใช้สอบกับผู้สอบที่มีความสามารถ แตกต่างกันนั้นจะถูกเลือกให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน มากกว่าคิดจากค่า เฉลี่ยความสามารถของกลุ่มผู้สอบ (Weiss, 1974: 1) โดยการทดสอบเทเลอร์ (Tailored Testing) มุ่งที่จะเลือกข้อสอบที่มีค่าความยากเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน ดังนั้นในการสอบครั้งหนึ่ง ๆ ผู้สอบแต่ละคนไม่จำเป็นต้องตอบข้อสอบเหมือนกันทุกข้อทั้งฉบับ แต่ ขึ้นอยู่กับผลการตอบในแบบทดสอบของแต่ละคน (Weiss and Kingsbury, 1984:361) ซึ่ง ผลจากการได้ตอบข้อสอบที่เหมาะสมกับระดับความสามารถของตนแล้ว ก็จะทำให้ทราบระดับความ สามารถที่แท้จริงของนักเรียนในเรื่องนั้น ๆ อยู่ในระดับใด

แนวคิดและวิธีการจัดรูปแบบการทดสอบเทเลอรัมืออยู่หลายวิธี ซึ่งในระยะแรกนั้นจะเสนอรูปแบบในการจัดข้อสอบไว้แล้วให้ผู้สอบสอบตามหลักการที่จัดไว้เป็นชุด ๆ เพื่อให้เหมาะสมกับความสามารถ ทั้งนี้เพื่อหลีกเลี่ยงการประมาณความสามารถของผู้สอบจากผลการตอบแต่ละข้อ (Weiss and Kingsbury, 1984: 361) ในระยะหลังในประเทศสหรัฐอเมริกาได้นำคอมพิวเตอร์ได้เข้ามาใช้ในทางการศึกษามากขึ้นจึงมีการประยุกต์การทดสอบเทเลอรัมือเข้าไปใช้กับคอมพิวเตอร์ โดยผู้สอบทำข้อสอบทีละข้อ ข้อคำถามที่ถูกเลือกให้สอบจะอยู่บนพื้นฐานการตอบของผู้สอบ ผลการตอบข้อสอบจะเป็นตัวประมาณระดับความสามารถ เพื่อคัดเลือกข้อใหม่ที่มีความยากเหมาะสม ซึ่งวิธีการนี้นำมาประยุกต์กับปัญหาทางการศึกษา 3 ประการ คือ 1) การทดสอบความรู้เพื่อจำแนกนักเรียนว่ามีความรอบรู้ในเนื้อหานั้นหรือไม่ 2) การกำหนดเกรดให้นักเรียน และ 3) เพื่อประมาณความก้าวหน้าในการเรียน (Weiss and Kingsbury, 1984:361) จากผลการศึกษาของเบตซ์และไวส์ (Betz and Weiss, 1973 cited in Larkin and Weiss, 1975:1) พบว่าการทดสอบเทเลอรัมือใช้เวลาในการทดสอบน้อยลง เมื่อเปรียบเทียบกับทดสอบแบบดั้งเดิม โดยที่ความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบไม่ลดลง นอกจากนี้การทดสอบแบบเทเลอรัมือยังมีประโยชน์คือ การทดสอบมีมาตรฐานจัดความลำเอียงจากผู้ดำเนินการสอบ ผู้สอบจะได้สอบข้อสอบต่างชุดกันตามระดับความสามารถ (Urry, 1977:194)

อย่างไรก็ตามจากการที่ผู้วิจัยได้ศึกษาแบบทดสอบเทเลอรัมือชนิดต่าง ๆ เห็นว่า รูปแบบบางชนิดของแบบทดสอบเทเลอรัมือสามารถที่จะนำมาประยุกต์ใช้ได้อย่างเหมาะสมกับสภาพการวัดผลที่ปฏิบัติกันอยู่ และเนื่องจากปัญหาของการทดสอบแบบอิงเกณฑ์นั้น ยิ่งก่อให้เกิดความคลาดเคลื่อนของการตัดสินทั้งนี้ เพราะไม่สามารถค้นหาระดับความสามารถที่ใกล้เคียงกับความสามารถจริงของผู้สอบได้ จึงมีผลกระทบต่อความเที่ยงตรงของการตัดสิน

จากปัญหาและความคลาดเคลื่อนในการจำแนกการรอบรู้ของการทดสอบแบบเดิม ดังที่กล่าวมาแล้ว เครื่องมือวัดหรือแบบสอบนับว่ามีความสำคัญอย่างมาก ซึ่งเมื่อเราพูดถึงเครื่องมือเราก็มักจะพูดถึงเรื่องของความเที่ยงตรงและความเชื่อมั่นว่ามีมากน้อยขนาดไหน แต่สิ่งที่ยังไม่

ค่อยมีการพูดถึงกันก็คือ ความเที่ยงตรงในการจำแนก (Classification Validity) ทั้งนี้ในการเรียนแบบผ่านเกณฑ์หรือแบบรอบรู้ (Mastery Learning) นั้น ผลจากการสอบในจุดประสงค์หนึ่ง ๆ จะต้องมีการตัดสินหรือจำแนกความรอบรู้ ซึ่งถือว่าเป็นขั้นตอนที่มีความสำคัญยิ่งหลังจากการดำเนินการสอบมาแล้วทั้งหมด ดังนั้น หากเครื่องมือวัดหรือแบบสอบขาดคุณสมบัติในสิ่งนี้จะทำให้การจำแนก หรือตัดสินผิดพลาดได้ ซึ่งแบบทดสอบอิงเกณฑ์ก็ยังมีจุดอ่อนในสิ่งนี้อยู่ ดังนั้น ผู้วิจัยจึงสนใจที่จะประยุกต์ใช้แบบทดสอบเทเลอร์เพื่อใช้ในการจำแนกความรอบรู้ ทั้งนี้เพราะว่าเป็นการทดสอบที่สามารถค้นหาระดับความสามารถที่ใกล้เคียงกับระดับความสามารถที่แท้จริงของผู้สอบได้ ดังนั้นจึงน่าจะให้ความเที่ยงตรงในการตัดสินความรอบรู้สูงกว่าการทดสอบแบบเดิม โดยผู้วิจัยเลือกรูปแบบการทดสอบตามที่ไรท์และสโตน (Wright and Stone, 1979 อ้างถึงใน สุธรรม์ จันทน์หอม 2529 หน้า 11) เสนอ ซึ่งมีวิธีการทดสอบเทเลอร์ 3 รูปแบบคือ 1) การทดสอบตามสภาพ (Status Tailoring) 2) การให้ทดลองทำ (Performance Tailoring) 3) การให้ตัดสินใจเลือกเอง (Self-Tailoring) ซึ่งทั้ง 3 รูปแบบเป็นวิธีการที่ง่าย ไม่ยุ่งยากซับซ้อนอีกทั้งมีวิธีการประมาณค่าผลการตอบด้วยมือ ซึ่งสามารถนำไปใช้ในระดับประถมศึกษาได้ พร้อมทั้งรูปแบบการทดสอบเทเลอร์แบบแตกแขนง (Branched Test) ซึ่งเสนอรูปแบบโดย ฟิชเชอร์และเพนเดล (Fischer and Pendl, 1980) ซึ่งก็เป็นรูปแบบที่ใกล้เคียงกับที่ไรท์และสโตน (Wright and Stone) โดยผู้วิจัยจะศึกษาผลการใช้แบบทดสอบอิงเกณฑ์ที่เพิ่มจำนวนข้อสอบเป็น 10 ข้อ 20 ข้อ และ 30 ข้อ และแบบทดสอบเทเลอร์ทั้ง 4 รูปแบบว่าจะให้ความเที่ยงตรงในการจำแนกความรอบรู้ความถูกต้องในการจำแนกความรอบรู้ ระดับความสามารถ ค่าสาระประโยชน์ของข้อสอบ และค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถแตกต่างกันอย่างไร

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีความมุ่งหมายของการวิจัย คือ

1. เพื่อศึกษาความเที่ยงตรงในการจำแนกความรอบรู้ จากการใช้แบบทดสอบอิงเกณฑ์ 10, 20 และ 30 ข้อ แบบทดสอบเทเลอร์แบบแตกแขนง (Branched Test) แบบทดสอบ

เทเลอร์แบบทดสอบตามสภาพ (Status Tailoring) แบบทดสอบเทเลอร์แบบทดลองทำ (Performance Tailoring) และแบบทดสอบเทเลอร์แบบตัดสินใจเลือกเอง (Self-Tailoring) เมื่อใช้เกณฑ์การตัดสินใจโดยผู้เชี่ยวชาญ

2. เพื่อเปรียบเทียบความถูกต้องของการจำแนกความรู้ จากการทดสอบแบบอิงเกณฑ์ และการทดสอบเทเลอร์ทั้ง 4 รูปแบบ
3. เพื่อเปรียบเทียบค่าประมาณความสามารถที่ได้จากผลการตอบ จากการทดสอบแบบอิงเกณฑ์และการทดสอบเทเลอร์ทั้ง 4 รูปแบบ
4. เพื่อเปรียบเทียบค่าสาระประโยชน์ของข้อสอบ (Item Information Function) ของการทดสอบแบบอิงเกณฑ์และการทดสอบเทเลอร์ทั้ง 4 รูปแบบ
5. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถจากผลการสอบ การทดสอบแบบอิงเกณฑ์ และการทดสอบเทเลอร์ทั้ง 4 รูปแบบ

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบว่าการทดสอบแบบอิงเกณฑ์ การทดสอบเทเลอร์แบบแตกแขนง การทดสอบเทเลอร์แบบทดสอบตามสภาพ การทดสอบเทเลอร์แบบทดลองทำ และการทดสอบเทเลอร์แบบตัดสินใจเลือกเอง มีความเที่ยงตรงในการจำแนกความรู้ของนักเรียนได้หรือไม่ เพียงไร
2. ทำให้ทราบว่าการจำแนกความรู้จากการทดสอบแบบอิงเกณฑ์และการทดสอบเทเลอร์ทั้ง 4 รูปแบบมีความแตกต่างกันหรือไม่
3. ทำให้ทราบค่าประมาณความสามารถที่ได้จากผลการทดสอบแบบอิงเกณฑ์และการทดสอบแบบเทเลอร์ทั้ง 4 รูปแบบแตกต่างกันมากน้อยเพียงใด
4. ทำให้ทราบค่าสาระประโยชน์ของข้อสอบจากการทดสอบแบบอิงเกณฑ์ และการทดสอบแบบเทเลอร์ทั้ง 4 รูปแบบแตกต่างกันหรือไม่เพียงไร
5. ทำให้ทราบความคลาดเคลื่อนของค่าความสามารถจากการตอบแบบทดสอบอิงเกณฑ์ และแบบทดสอบเทเลอร์แตกต่างกันหรือไม่เพียงไร
6. เป็นแนวทางในการเลือกใช้แบบทดสอบเทเลอร์ในการเรียนการสอนต่อไป

7. เป็นแนวทางในการส่งเสริมให้มีการพัฒนาการทดสอบแบบเทเลเลอร์เพื่อการประเมินผลสัมฤทธิ์ทางการเรียน

8. เป็นแนวทางในการศึกษาวิจัยการทดสอบแบบเทเลเลอร์ต่อไป

สมมติฐานการวิจัย

1. แบบทดสอบแบบอิงเกณฑ์ แบบทดสอบเทเลเลอร์แบบแตกแขนง (Branched Test) แบบทดสอบเทเลเลอร์แบบทดสอบตามสภาพ (Status Tailoring) แบบทดสอบเทเลเลอร์แบบทดลองทำ (Performance Tailoring) และแบบทดสอบเทเลเลอร์แบบตัดสินใจเลือกเอง (Self-Tailoring) ให้ความเที่ยงตรงในการจำแนกความรู้

2. แบบทดสอบแบบอิงเกณฑ์และแบบทดสอบเทเลเลอร์ทั้ง 4 รูปแบบให้ผลในการจำแนกความรู้แตกต่างกัน

3. ค่าความสามารถที่ประมาณจากแบบทดสอบแบบอิงเกณฑ์ และแบบทดสอบเทเลเลอร์ทั้ง 4 รูปแบบ แตกต่างกัน

4. ค่าสาระประโยชน์ของข้อสอบ (Item Information Function) ของแบบทดสอบแบบอิงเกณฑ์ต่ำกว่าแบบทดสอบเทเลเลอร์ทั้ง 4 รูปแบบ โดยที่แบบทดสอบเทเลเลอร์ทั้ง 4 แบบให้ค่าสาระประโยชน์ของข้อสอบแตกต่างกัน

5. ค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถ จากการตอบแบบทดสอบแบบอิงเกณฑ์สูงกว่าแบบทดสอบเทเลเลอร์ทั้ง 4 รูปแบบ โดยที่แบบทดสอบเทเลเลอร์ทั้ง 4 รูปแบบให้ความคลาดเคลื่อนของการประมาณค่าความสามารถจากผลการสอบแตกต่างกัน

ข้อตกลงเบื้องต้น

1. ตามข้อตกลงเบื้องต้น (Assumption) ของทฤษฎีการตอบข้อคำถาม (Item Response Theory : IRT) ที่กำหนดไว้ในเรื่องแบบทดสอบเป็นมิติเดียว (Unidimension) ความเป็นอิสระที่ระดับความสามารถเดียวกัน (Local Independence) และโด่งลักษณะของ

ข้อคำถาม (Item Characteristic Curve) นั้น ในการสร้างข้อสอบที่ใช้ในการวิจัยจะไม่
 พิจารณาข้อตกลงทั้ง 3 ประการ แต่ผู้วิจัยจะใช้หลักเหตุผลทางด้านผู้เชี่ยวชาญ เพื่อให้ข้อสอบแต่
 ละข้อมีคุณสมบัติตามข้อตกลงเบื้องต้นของทฤษฎี

2. ในการกำหนดกลุ่มนักเรียนออกเป็นกลุ่มรอบรู้และไม่รอบรู้อย่างแท้จริง โดยการใช้การ
 ทดสอบวิธีเบย์ส (Bayesian Strategies) ค่าประมาณความสามารถที่ได้จากการทดสอบโดยวิธี
 เบย์ส ที่อยู่สูงหรือต่ำกว่าเกณฑ์การตัดสินความรอบรู้ โดยที่ช่วงความเชื่อมั่นของค่าประมาณความ
 สามารถไม่ครอบคลุมเกณฑ์ที่กำหนด จะถือว่าเป็นกลุ่มที่รอบรู้และไม่รอบรู้อย่างแท้จริงตามลำดับ

3. ในการทดสอบแบบอิงเกณฑ์และการทดสอบแบบเทเลอร์ (Tailored Testing)
 ทั้ง 4 รูปแบบนั้น ผลการทดสอบในแบบทดสอบแต่ละชนิดเป็นอิสระจากกัน คือ ผลการทดสอบ
 ในแบบทดสอบแบบหนึ่งไม่ส่งผลต่อการตอบในอีกแบบทดสอบหนึ่ง

4. ระดับความสามารถของนักเรียนกลุ่มตัวอย่างที่ใช้ในการวิจัย จำนวน 115 คน
 มีการกระจายโค้งปกติ

ขอบเขตของการวิจัย

การวิจัยครั้งนี้กำหนดขอบเขตไว้ดังนี้

1. ประชากรที่ใช้ในการศึกษาเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในภาคเรียนที่ 3
 ปีการศึกษา 2529 ของโรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดเชียงราย

2. เนื้อหาที่ใช้เป็นเนื้อหาในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จุดประสงค์การ
 เรียนรู้ข้อที่ 9 แก่โจทย์ปัญหาเกี่ยวกับทศนิยมได้

3. ตัวแปรที่ใช้ในการศึกษา มีดังนี้

3.1 ตัวแปรอิสระ ได้แก่

3.1.1 แบบทดสอบแบบอิงเกณฑ์ที่ใช้ข้อสอบจำนวน 10 ข้อ

3.1.2 แบบทดสอบแบบอิงเกณฑ์ที่ใช้ข้อสอบจำนวน 20 ข้อ

3.1.3 แบบทดสอบแบบอิงเกณฑ์ที่ใช้ข้อสอบจำนวน 30 ข้อ

3.1.4 แบบทดสอบเทเลอร์แบบแตกแขนง (Branched Test)

3.1.5 แบบทดสอบเทเลอร์แบบทดสอบตามสภาพ (Status Tailoring)

3.1.6 แบบทดสอบเทเลอร์แบบทดสอบการทำ (Performance Tailoring)

3.1.7 แบบทดสอบเทเลอร์แบบตัดสินใจเลือกทำเอง (Self-Tailoring)

3.2 ตัวแปรตาม ได้แก่

3.2.1 ความเที่ยงตรงในการจำแนกการรอบรู้

3.2.2 ความถูกต้องของการจำแนกการรอบรู้

3.2.3 ค่าประมาณความสามารถ จากผลการตอบแบบทดสอบแบบอิงเกณฑ์ (10, 20 และ 30 ข้อ) และแบบทดสอบแบบเทเลอร์ทั้ง 4 แบบ

3.2.4 ค่าสาระประโยชน์ของข้อสอบ (Item Information Function)

3.2.5 ค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถจากผลการทดสอบแบบอิงเกณฑ์ (10, 20 และ 30 ข้อ) และแบบทดสอบแบบเทเลอร์ทั้ง 4 รูปแบบ

4. เกณฑ์ที่ใช้ในการจำแนกการรอบรู้ไม่รอบรู้ จากผลการสอบในแต่ละรูปแบบของการทดสอบนั้น กำหนดโดยให้ผู้เชี่ยวชาญระบุข้อสอบในกลุ่มข้อสอบที่นักเรียนจะต้องทำถูกเป็นอย่างต่ำ จึงจะถือว่าเป็นผู้รอบรู้ในจุดประสงค์

5. การจำแนกนักเรียนเป็นกลุ่มรอบรู้ และไม่รอบรู้อย่างแท้จริงนั้น ผู้วิจัยจะใช้การทดสอบโดยวิธีเบส์ (Bayesian Strategies)

นิยามศัพท์เฉพาะ

1. แบบทดสอบแบบอิงเกณฑ์ หมายถึง แบบทดสอบวัดความสามารถแก้โจทย์ปัญหา ทศนิยมในระดับชั้นประถมศึกษาปีที่ 6 ที่ผู้วิจัยสร้างขึ้น เมื่อนำผลการสอบแต่ละคนมาตัดสินโดยการ ใช้จุดตัดที่กำหนดขึ้นเป็นเกณฑ์สำคัญในการแปลความหมาย (Interpretation) ของคะแนนการ

สอบ (Brennan, 1984 อ้างถึงในสุพันธุ์ สุกมลสันต์ 2528 หน้า 19) ซึ่งเป็นลักษณะเฉพาะของแบบทดสอบชนิดนี้มากกว่าคุณลักษณะอื่น เช่น การสร้างข้อสอบและการเลือกข้อสอบเพราะลักษณะดังกล่าวของข้อสอบอิงเกณฑ์ และแบบทดสอบอิงกลุ่มไม่แตกต่างกันอย่างเด่นชัด (Block and Dockrell, 1984 อ้างถึงในสุพันธุ์ สุกมลสันต์ 2528 หน้า 19)

2. แบบทดสอบเทเลอร์ (Tailored Testing) หมายถึง แบบทดสอบที่มุ่งจะเลือกข้อสอบที่มีระดับความยากเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน โดยที่การสอบครั้งหนึ่ง ผู้สอบแต่ละคนอาจไม่จำเป็นต้องตอบข้อสอบเหมือนกันทุกข้อ แต่ขึ้นอยู่กับผลการตอบข้อสอบในแบบทดสอบของแต่ละคน (Weiss and Kingsbury, 1984:361) ซึ่งในการวิจัยครั้งนี้ได้แก่

2.1 แบบทดสอบเทเลอร์แบบแตกแขนง (Branched Test) หมายถึง แบบทดสอบที่จัดแบบทดสอบไว้ตามรูปแบบของนิชเชอร์และเพนเดิล (Fiseher and Pendl, 1980) เพื่อใช้สอบให้เหมาะสมกับระดับความสามารถของนักเรียน ประกอบด้วยข้อสอบ 3 ชุด ชุดแรกมี 1 ฉบับ ชุดที่ 2 มี 5 ฉบับ และชุดที่ 3 มี 9 ฉบับ แต่ละฉบับมี 10 ข้อ ผลจากการสอบชุดแรกจะเป็นเครื่องชี้การเลือกแบบทดสอบในชุดที่ 2 1 ฉบับ และผลคะแนนตอบรวมจากชุดที่ 1 และ 2 จะเป็นเครื่องชี้เพื่อเลือกแบบทดสอบชุดที่ 3 อีก 1 ฉบับ

2.2 แบบทดสอบเทเลอร์แบบทดสอบตามสภาพ (Status Tailoring) หมายถึง แบบทดสอบซึ่งครูผู้สอนกลุ่มวิชาทักษะคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จะเป็นผู้กำหนดหรือเลือกแบบทดสอบที่เหมาะสมกับความสามารถของนักเรียนแต่ละคน โดยพิจารณาจากข้อมูลผลการเรียนกลุ่มวิชาทักษะคณิตศาสตร์ในระดับชั้นประถมศึกษาปีที่ 6

2.3 แบบทดสอบเทเลอร์แบบให้ทดลองทำ (Performance Tailoring) หมายถึง แบบทดสอบที่ให้นักเรียนทดลองทำข้อสอบจำนวนหนึ่งหรือแบบทดสอบนำร่อง (Pilot Test) ผลจากการทำจะประมาณค่าความสามารถและความคลาดเคลื่อน เพื่อจัดข้อสอบใหม่ให้ทำโดยที่ค่าความยากของข้อสอบอยู่ในช่วงของค่าความสามารถ $\pm$ ค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณความสามารถ

2.4 แบบทดสอบเทเลอร์แบบให้ตัดสินใจเลือกทำเอง (Self - Tailoring)

หมายถึง แบบทดสอบเทเลอร์ที่ให้นักเรียนเลือกจุดเริ่มต้นในการทำข้อแรกเอง โดยข้อสอบในแบบทดสอบได้จัดเรียงความยากไว้ ค่าความยากของข้อสอบข้อแรกที่เลือกเป็นจุดเริ่มต้นนั้นจะต้องไม่ยากหรือไม่ง่ายเกินไปที่จะใช้ความสามารถที่มีอยู่ตอบ เฉลยผลการตอบทีละข้อและให้นักเรียนทำต่อเนื่องไปเรื่อย ๆ จะหยุดทำข้อสอบก็ต่อเมื่อนักเรียนตอบผิดติดต่อกัน 3 ข้อ

3. ความยากของข้อสอบ หมายถึง ค่าความสามารถตรงจุดเปลี่ยนโค้งคุณลักษณะของข้อสอบแต่ละข้อ (Item Characteristic Curve) ซึ่งเป็นระดับความสามารถที่จะตอบข้อสอบนั้นได้ถูกต้องด้วยค่าความน่าจะเป็น .50 เมื่อไม่มีการเดา (ผจงจิต อินทสุวรรณ 2528 หน้า 16) การวิจัยครั้งนี้ทำโดยใช้โมเดลของราชคซึ่งใช้พารามิเตอร์ 1 ตัวด้วยโปรแกรมไบคาล 3 (BICAL 3)

4. เกณฑ์การตัดสินความรอบรู้ หมายถึง ระดับความสามารถที่ใช้เป็นเกณฑ์มาตรฐานขั้นต่ำสุดที่จะจำแนกหรือตัดสินนักเรียนว่าเป็นผู้รอบรู้และไม่รอบรู้ จากผลการทดสอบแบบอิงเกณฑ์และการทดสอบเทเลอร์แบบแตกแขนง (Branched Test) การทดสอบเทเลอร์แบบทดสอบตามสภาพ (Status Tailoring) การทดสอบเทเลอร์แบบตัดสินใจเลือกทำเอง (Self - Tailoring) โดยเกณฑ์นี้กำหนดโดยให้ผู้เชี่ยวชาญเป็นผู้กำหนดจากกลุ่มข้อสอบ

5. ความเที่ยงตรงในการจำแนกความรอบรู้ หมายถึง ความสอดคล้องระหว่างผลการจำแนกหรือตัดสินนักเรียน โดยใช้ข้อมูลจากผลการทดสอบว่าเป็นผู้รอบรู้หรือไม่รอบรู้กับสถานะการรอบรู้หรือไม่รอบรู้อย่างแท้จริงของนักเรียนแต่ละคนว่ามีมากน้อยเพียงใด สำหรับการวิจัยครั้งนี้หาได้จากสหสัมพันธ์ระหว่างสถานะการรอบรู้ที่แท้จริงของนักเรียน (student's actual mastery states) และสถานะการรอบรู้ที่ประมาณการ (estimated mastery states) จากการทดสอบ ซึ่งหาโดยสหสัมพันธ์แบบฟาย (phi correlation)

6. ค่าความสามารถ หมายถึง ค่าประมาณระดับความสามารถทางการเรียนของนักเรียนแต่ละคน ซึ่งได้จากการนำผลการทดสอบแบบอิงเกณฑ์ และแบบทดสอบเทเลอร์แบบแตก

แผนการทดสอบเทเลอร์แบบทดสอบตามสภาพ การทดสอบเทเลอร์แบบทดลองทำ และการทดสอบเทเลอร์แบบตัดสินใจเลือกทำเอง ไปประมาณค่าความสามารถตามโมเดลราสค์ วิธี UFORM

7. ค่าสาระประโยชน์ของข้อสอบ (Item Information Function) หมายถึง ค่าสัดส่วนกลับกำลังสองของช่วงแห่งความเชื่อมั่นที่เป็นเอซิมโทติก (asymptotic confidence interval) ในการประมาณค่าความสามารถของผู้สอบจากผลการตอบของเขา (ผจงจิต อินทสุวรรณ 2528 หน้า 97) ซึ่งค่านี้เป็นเครื่องแสดงถึงความถูกต้องแน่นอน (precision) ของค่าที่ประมาณได้

8. ค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถการทดสอบ หมายถึง ค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณความสามารถจากผลการทดสอบแบบอิงเกณฑ์ และการทดสอบเทเลอร์ทั้ง 4 รูปแบบ ซึ่งคำนวณหาโดยวิธี UFORM

9. โมเดลราสค์ (Rasch Model) หมายถึง วิธีการวิเคราะห์ข้อสอบหลายข้อ และวิธีการจัดคุณลักษณะแฝงของการสอบวิธีหนึ่งที่มีสมมติฐานเบื้องต้นว่าโอกาสที่บุคคลจะทำข้อสอบข้อใดได้ถูกต้องหรือไม่ ขึ้นอยู่กับความสามารถของตนเอง (θ) และระดับความยากง่ายของข้อสอบ (β) นั้นเท่านั้น กล่าวคือ (Hambleton and Cook; 1977 : 82)

$$P_{\theta}(\theta) = \frac{\exp(\theta - b_{\theta})}{1 + \exp(\theta - b_{\theta})}$$

ในเมื่อ $\exp = 2.7182818$ และ $b = \beta$