



บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัด เป็นกิจกรรมที่สำคัญในการแสวงหาความรู้ในศาสตร์ต่าง ๆ ตัวแปรในศาสตร์ใด สามารถวัดได้ละเอียดถี่ถ้วน ถูกต้องแม่นยำ ก็จะทำให้การวิจัยและการศึกษาหาความรู้ในศาสตร์นั้นทำได้สะดวกและสามารถพัฒนาให้ก้าวหน้าไปได้อย่างรวดเร็ว ข้อความรู้นั้นก็จะถูกต้องสมบูรณ์สามารถนำไปประยุกต์ใช้กับสถานการณ์ต่าง ๆ ได้อย่างกว้างขวางมากกว่าศาสตร์ที่การวัดมีความถูกต้องแม่นยำในระดับที่น้อยกว่า หน้าที่อันหนึ่งที่นักวิชาการในศาสตร์ต่าง ๆ นิยมกระทำคือ พยายามพัฒนากลวิธีการวัดตัวแปรในศาสตร์ของตนให้มีประสิทธิภาพมากที่สุดเท่าที่จะทำได้

วิธีการและเครื่องมือที่ใช้วัดตัวแปรทางสังคมศาสตร์นั้น เป็นที่ยอมรับกันว่ามีประสิทธิภาพค่อนข้างต่ำ เมื่อเทียบกับวิธีการและเครื่องมือที่ใช้วัดตัวแปรทางวิทยาศาสตร์ธรรมชาติ แต่นักวิชาการทางด้านสังคมศาสตร์ก็ได้ทุ่มเทความพยายามที่จะพัฒนากลวิธีการวัดในศาสตร์ของตนให้มีประสิทธิภาพสูงขึ้น ด้วยการคิดหากลวิธีใหม่ ๆ ในการวัดหรือพัฒนาแนวคิดทฤษฎีที่จะนำมาใช้เป็นแนวทางศึกษาเพื่อหาทางปรับปรุงเครื่องมือวัดให้ดียิ่งขึ้น ในวงการการศึกษาซึ่งเป็นสาขาหนึ่งของสังคมศาสตร์ก็ได้มีความพยายามที่จะพัฒนาเครื่องมือวัดตัวแปรทางการศึกษาให้มีประสิทธิภาพอยู่อย่างต่อเนื่องเสมอมาโดยตลอด

ผลสัมฤทธิ์ทางการเรียนถือได้ว่าเป็นตัวแปรหลักที่สำคัญในศึกษาศาสตร์ เป็นตัวแปรที่นักการศึกษาให้ความสนใจอย่างมากมาโดยตลอดซึ่งอายุของศึกษาศาสตร์ และคาดว่ายังจะเป็นตัวแปรที่ยังอยู่ในความสนใจของนักการศึกษาต่อไปอีกนานในอนาคต ตลอดระยะเวลาที่ผ่านมา นักการศึกษาได้พยายามพัฒนากลวิธีวัดตัวแปรผลสัมฤทธิ์ทางการเรียนขึ้นมามากมายหลายวิธี แต่กลวิธีที่ค่อนข้างจะมีบทบาทมากก็คือ การสอบ เพราะเป็นกลวิธีที่นิยมใช้กันอย่างแพร่หลาย ในบรรดาการสอบวิธีต่าง ๆ นั้น การสอบข้อเขียนนับว่ามีบทบาทสำคัญในการสอบวัดผลสัมฤทธิ์ทางการเรียนมาโดยตลอด และคาดว่ายังจะมีบทบาทสำคัญต่อไปอีกในอนาคต

แต่เดิมนั้นการสอบข้อเขียนนิยมใช้ข้อสอบแบบความเรียง (essay type) หรือแบบ
 อัตนัย ต่อมาก็พบว่า การตอบข้อสอบแบบความเรียงนั้นผู้ตอบต้องใช้ความรู้ความสามารถหลาย ๆ
 อย่างพร้อมกัน จนไม่สามารถระบุได้อย่างเฉพาะเจาะจงว่าข้อสอบข้อหนึ่ง ๆ วัดความสามารถด้าน
 ใด ประกอบกับมีปัญหาในเรื่องของการตรวจให้คะแนนที่มีความเป็นปรนัยน้อย เป็นผลให้ผลการ
 สอบขาดความเที่ยงและความตรงตามมา อีกทั้งข้อสอบแบบนี้ต้องใช้เวลาในการตอบมากทำให้
 ถ้ามคำถามได้น้อยข้อภายในเวลาที่กำหนด ทำให้ไม่สามารถสอบวัดได้ครอบคลุมเนื้อหาที่ต้องการ
 วัดได้ครบถ้วน และในกรณีที่ผู้เข้าสอบเป็นจำนวนมากแต่มีเวลาในการตรวจให้คะแนนไม่มากทำ
 ให้ไม่สามารถรายงานผลการสอบได้รวดเร็วทันกับความต้องการใช้ผลการสอบได้ ปัจจัยเหล่านี้
 ทำให้การสอบวัดโดยข้อสอบแบบความเรียงมีปัญหา ไม่มีความสะดวกในการปฏิบัติเท่าที่ควร
 เพื่อแก้ปัญหาของการใช้ข้อสอบแบบความเรียงนักวัดผลการศึกษาจึงได้เสนอให้มีการใช้ข้อสอบแบบ
 ปรนัย (objective type) ซึ่งเป็นข้อสอบที่มีคำถามที่เฉพาะเจาะจงมากขึ้น สามารถตัดสิน
 ความถูกต้องของคำตอบได้ง่าย และสามารถระบุได้ค่อนข้างแน่ชัดว่าในการตอบข้อสอบแต่ละข้อนั้น
 ผู้ตอบจะต้องใช้ความรู้ความสามารถด้านใด การตอบคำถามแต่ละข้อใช้เวลาไม่มากนักจึงใช้ข้อสอบ
 ถ้ามได้หลายข้อ ทำให้สามารถถ้ามได้ครอบคลุมเนื้อหาที่ต้องการวัดได้มาก ประกอบกับการตรวจ
 ให้คะแนนมีความเป็นปรนัยสูง ทำได้รวดเร็ว สามารถรายงานผลการสอบได้ทันกับความต้องการ
 ใช้แม้จะมีผู้เข้าสอบเป็นจำนวนมากก็ตาม ข้อสอบแบบปรนัยจึงเป็นที่นิยมใช้กันแพร่หลายมากใน
 ปัจจุบัน

ข้อสอบแบบปรนัยที่ใช้กันอยู่ขณะนี้รูปแบบที่นิยมใช้กันอยู่ 5 รูปแบบ คือ แบบตอบสั้น
 (short answer) แบบเติมข้อความ (completion) แบบถูกผิด (true-false) แบบจับคู่
 (matching) และแบบเลือกตอบ (multiple-choice) ข้อสอบแต่ละรูปแบบก็มีข้อดีและข้อจำกัด
 ในแง่ต่าง ๆ แตกต่างกันไป ในบรรดาข้อสอบปรนัยทั้ง 5 รูปแบบนี้ ข้อสอบแบบเลือกตอบเป็น
 รูปแบบที่นิยมใช้กันแพร่หลายกว่ารูปแบบอื่นๆ ด้วยเหตุผลที่สำคัญคือ เป็นรูปแบบที่เอื้อต่อการวัด
 ความรู้ความสามารถในระดับสูงได้ สามารถถ้ามได้ครอบคลุมเนื้อหาสาระที่ต้องการวัดได้มาก
 การตรวจให้คะแนนทำได้ง่าย รวดเร็ว และมีความเป็นปรนัยสูง ข้อมูลที่ได้จากการตอบข้อสอบ
 รูปแบบนี้สามารถนำไปวิเคราะห์ด้วยวิธีการทางสถิติได้หลายรูปแบบ นำไปสู่การนวัตนาทฤษฎี
 เกี่ยวกับการวัดให้ก้าวหน้า ถึงกับมีผู้กล่าวว่า ถ้าไม่มีข้อสอบแบบเลือกตอบแล้วทฤษฎีการวัดสมัยใหม่
 คงไม่อาจพัฒนาก้าวหน้ามาอย่างที่เป็นอยู่นี้ได้ (Wood 1977: 193)

ข้อสอบเลือกตอบชนิดแบบฉบับ

ข้อสอบแบบเลือกตอบชนิดที่นิยมใช้กันแพร่หลายอยู่ในขณะนี้ เป็นข้อสอบชนิดที่กำหนดให้ตัวเลือก 1 ตัว เป็นคำตอบที่ถูกต้อง ตัวเลือกที่เหลืออื่น ๆ เป็นตัวลวง ผู้ตอบจะต้องเลือกคำตอบที่ถูกต้องออกมาจากตัวเลือกที่กำหนดให้ โดยทำเครื่องหมายที่ตัวเลือกที่ตนเลือกเพียงตัวเลือกเดียว ถ้าเลือกได้ถูกต้องก็จะได้คะแนน แต่ถ้าเลือกไม่ถูกต้องก็จะได้คะแนน ข้อสอบแบบนี้ต่อไปนี้จะขอเรียกว่า ข้อสอบเลือกตอบชนิดแบบฉบับ (typical multiple-choice, T.MC) แม้ข้อสอบเลือกตอบชนิดแบบฉบับจะมีข้อดีหลายประการดังกล่าวแล้ว แต่ก็มีข้อบกพร่องที่ได้รับการวิพากษ์วิจารณ์อยู่ไม่น้อย เช่น เปิดโอกาสให้มีการตอบถูกต้องโดยการเดา วัดความสามารถในระดับที่ลึกซึ้งได้ยาก ผู้ตอบไม่มีโอกาสได้แสดงความสามารถในการเขียนเรียบเรียงคำตอบ ซึ่งก็เป็นจุดอ่อนของข้อสอบปรนัยทุกรูปแบบและเป็นที่ทราบกันโดยทั่วไปในหมู่ผู้ใช้ข้อสอบแบบปรนัย แต่ยังมีข้อบกพร่องของข้อสอบเลือกตอบชนิดแบบฉบับที่นักวัดผลตระหนักว่าเป็นสาเหตุทำให้ข้อสอบเลือกตอบชนิดนี้วัดได้ถูกต้องน้อยลงกว่าที่ควรจะเป็น ได้แก่

ประการแรก การให้มีตัวเลือกที่เป็นคำตอบที่ถูกต้องเพียงตัวเดียวอาจไม่เหมาะสมในกรณีที่ข้อความรู้ที่นำมาสร้างเป็นปัญหานั้นมีคำตอบได้หลายคำตอบและเป็นคำตอบที่ถูกต้องได้เท่า ๆ กัน การจะตัดสินว่าคำตอบใดเป็นคำตอบที่ถูกต้องที่สุดหรือดีที่สุดเพียงคำตอบเดียวบางทีไม่อาจทำได้ การเขียนข้อสอบให้มีคำตอบที่ถูกต้องเพียงคำตอบเดียวจึงเป็นเรื่องยากมาก ถ้าจะแก้ปัญหารื้อนี้ โดยการแยกเอาคำตอบที่ถูกต้องคำตอบใดคำตอบหนึ่งมาเป็นตัวเลือกที่ผูกพันกับตัวเลือกที่ผิดอื่น ๆ ก็ไม่อาจวัดความรู้ที่สมบูรณ์ครบถ้วนในเรื่องนั้นได้ เพราะถ้าผู้ตอบเลือกคำตอบที่ถูกต้องได้ก็ไม่ได้หมายความว่า เขาจะทราบคำตอบที่ถูกต้องอื่น ๆ ที่เราไม่ได้นำมาเป็นตัวเลือกได้ ถ้าต้องการทราบทั้งหมดก็คงจะต้องตั้งคำถามเดิมซ้ำอีกและใช้คำตอบที่เป็นไปได้อื่น ๆ มาเป็นตัวเลือกที่ถูกต้องใหม่ ก็จะเป็นการถามซ้ำและทำให้ข้อสอบมีจำนวนข้อมากขึ้นโดยไม่จำเป็น ปัญหาการมีคำตอบถูกต้องหลายคำตอบ เรามักจะพบเสมอในหลายสาขาวิชาและเราจำเป็นต้องทราบว่า ผู้ตอบรู้คำตอบที่ถูกต้องทั้งหมดของปัญหานั้นหรือไม่ ในกรณีเช่นนี้ข้อสอบเลือกตอบชนิดแบบฉบับไม่อาจทำหน้าที่ได้ดีเท่าที่ควร

ประการที่สอง การมีคำตอบที่ถูกต้องเพียงคำตอบเดียวอาจจะเป็นการช่วยให้ผู้ตอบหาคำตอบที่ถูกต้องได้ง่ายขึ้น ยิ่งถ้าเป็นผู้ที่มึนประสบการณ์ในการตอบข้อสอบหรือเป็นผู้ที่มีความรู้อยู่บ้าง

แต่ไม่สมบูรณ์ครบถ้วนอาจตอบข้อสอบถูกได้ไม่ยากนัก เช่นอาจจะตอบโดยการค่อย ๆ กำจัดตัวเลือกที่ผิดออกไป ซึ่งก็ทำได้ง่ายเพราะโอกาสที่ตัวเลือกต่าง ๆ จะเป็นตัวเลือกที่ผิดนั้นมีมาก หรือบางครั้งเมื่อผู้ตอบเลือกตัวเลือกที่เขาคิดว่าถูกได้แล้วก็จะไม่พิจารณาตัวเลือกอื่น ๆ โดยจะถือว่าตัวเลือกอื่น ๆ ที่เหลือนั้นผิดทันที ทั้ง ๆ ที่เขาอาจไม่ทราบว่าตัวเลือกเหล่านั้นบางตัวเลือกทำไม่ถึงผิด ซึ่งถ้าเขาไม่ทราบมาก่อนว่าข้อสอบนี้เมื่อคำตอบที่ถูกต้องเพียงคำตอบเดียวเขาอาจจะเลือกตัวเลือกอื่นที่เขาไม่ทราบว่าทำไมจึงผิดนั้นเป็นคำตอบที่ถูกอีกก็ได้ ดังที่ Kolstad และคณะ (1985: 27-31) ได้ทดลองให้ผู้ตอบข้อสอบ 2 กลุ่ม ตอบข้อสอบเลือกตอบที่มีคำตอบที่ถูกต้องเพียงคำตอบเดียว โดยกลุ่มหนึ่งทราบว่า เป็นข้อสอบเลือกตอบชนิดแบบฉบับที่มีคำตอบที่ถูกต้องเพียงคำตอบเดียว อีกกลุ่มหนึ่งก็ให้ตอบข้อสอบชุดเดียวกัน แต่ให้ตอบรับตัวเลือกที่ถูกและตอบปฏิเสธตัวเลือกที่ผิดโดยไม่ทราบว่า มีตัวเลือกที่เป็นคำตอบที่ถูกก็ตัว ผลปรากฏว่า กลุ่มที่ไม่ทราบว่าข้อสอบมีคำตอบที่ถูกเพียงคำตอบเดียวยอมรับคำตอบที่ผิดมากกว่ากลุ่มที่ทราบ ผลการทดลองนี้แสดงให้เห็นว่า ผู้ตอบข้อสอบเลือกตอบชนิดแบบฉบับส่วนใหญ่จะไม่เลือกคำตอบอื่นอีกนอกจากคำตอบที่เขาคิดว่าถูก แม้ว่าเขาจะไม่ทราบเหตุผลว่าทำไมคำตอบที่เป็นตัวจริงเป็นคำตอบที่ผิดก็ตาม นั่นคือผู้ตอบข้อสอบเลือกตอบชนิดแบบฉบับได้ถูกต้องนั้น ไม่จำเป็นต้องเป็นผู้รู้ เรื่องนั้นอย่างสมบูรณ์เสมอไป

ประการที่สาม การให้คะแนน 1 สำหรับผู้เลือกคำตอบได้ถูก และให้ 0 สำหรับผู้ที่ไม่เลือกคำตอบไม่ถูกนั้น ไม่สามารถแยกแยะผู้ที่มีความรู้ถูกต้องสมบูรณ์ ผู้ที่มีความรู้บางส่วน กับผู้ที่มีความรู้ไม่ถูกต้องออกจากกันได้เลย (Coombs, Milholand, and Womer 1965: 13-14) ความจริงแล้วในการตอบข้อสอบเลือกนั้น ผู้ตอบอาจจะมีความรู้ความสามารถในระดับที่แตกต่างกันอยู่ ดังที่ Little (1962 cited in Hughes, and Trimble 1965: 117) ได้วิเคราะห์ถึงระดับความรู้ของผู้ตอบที่นำไปสู่การตอบข้อสอบแต่ละข้อถูกหรือผิดว่ามีความเกี่ยวข้องกันอยู่ 6 ลักษณะ คือ

1. มีความรู้ที่ถูกต้องในเรื่องนั้นนำไปสู่การตอบถูก
2. มีความรู้บางส่วนนำไปสู่การตอบถูก
3. ไม่มีความรู้เลยเลือกคำตอบแบบเดาสุ่ม
4. มีความรู้บางส่วนนำไปสู่การตอบผิด
5. มีความรู้ที่ไม่ถูกต้องนำไปสู่การตอบผิด
6. รู้สึกว่าคำถามในข้อสอบไม่เหมาะสมและรู้คำตอบที่ผู้วัดต้องการให้ตอบ แต่เขาไม่เห็นด้วยและจะไม่ตอบคำตอบที่ถูกต้อง

และจากผลการวิจัยของ Coombs และคณะ (1965: 13-37) ก็พบว่ามีหลักฐานเพียงพอที่จะเชื่อได้ว่า ผู้ตอบข้อสอบเลือกตอบชนิดแบบฉบับส่วนหนึ่งมีความรู้บางส่วน ข้อสอบชนิดนี้จึงไม่มีความละเอียดถี่ถ้วนหรือไว (sensitive) พอที่จะจับความแตกต่างของระดับความรู้ที่ผู้เข้าสอบแต่ละคนมีอยู่ได้ เพราะมาตรวัดที่กำหนดไว้หยากกว่าที่ควรจะเป็น ทำให้ผลการวัดมีความถูกต้องน้อยลง

ถึงแม้ข้อสอบเลือกตอบชนิดแบบฉบับจะมีจุดอ่อนหลายข้อดังกล่าว แต่ข้อสอบปรนัยรูปแบบนี้ยังมีข้อได้เปรียบอยู่หลายประการที่มีคุณค่าพอที่จะยังคงใช้เป็นเครื่องมือในการวัดผลอยู่ นักวัดผลหลายท่านจึงได้พยายามหาทางแก้ไขจุดอ่อนเหล่านั้น เพื่อให้การใช้ข้อสอบแบบเลือกตอบมีประสิทธิภาพมากที่สุด วิธีการแก้ไขจุดอ่อนเหล่านี้พอสรุปได้เป็น 3 แนวทาง ได้แก่

1. **ปรับวิธีการตรวจให้คะแนน** วิธีการนี้ตัวข้อสอบและวิธีการตอบยังคงเหมือนกับข้อสอบเลือกชนิดแบบฉบับทุกประการ แต่จะมีการเปลี่ยนแปลงวิธีการตรวจให้คะแนน เช่น การให้คะแนนตามสูตร (scoring formula) หรือ การกำหนดน้ำหนักคะแนนให้กับตัวเลือกแต่ละตัวแตกต่างกัน เป็นต้น

2. **ปรับวิธีการตอบและการตรวจให้คะแนน** วิธีการนี้ผู้ตอบจะได้รับคำชี้แจงให้ตอบข้อสอบด้วยวิธีการที่แตกต่างจากการตอบข้อสอบเลือกตอบชนิดแบบฉบับ แต่ข้อสอบจะยังคงมีรูปแบบเหมือนเดิม คือมีคำตอบที่ถูกต้องเพียงตัวเลือกเดียว เช่น การให้ตอบปฏิเสธตัวเลือกที่ผิด การให้ตอบรับตัวเลือกที่ถูกและตอบปฏิเสธตัวเลือกที่ผิด การให้เลือกคำตอบได้หลายตัวเลือก การให้ผู้ตอบบอกระดับความมั่นใจในการตอบ เป็นต้น

3. **ปรับรูปแบบข้อสอบ วิธีการตอบและวิธีการตรวจให้คะแนน** วิธีการนี้ข้อสอบจะมีตัวเลือกที่เป็นคำตอบที่ถูกอยู่หลายตัวเลือก แต่วิธีการตอบและวิธีการตรวจให้คะแนนจะต่างไปจากข้อสอบเลือกตอบชนิดแบบฉบับ เช่น การใช้ข้อสอบเลือกตอบชนิดเชิงซ้อน (double or complex multiple-choice) ข้อสอบที่มีคำตอบถูกหลายตัวเลือกและให้เลือกตอบรับเฉพาะตัวเลือกที่ถูก (multiple-answer) ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก (multiple true-false) เป็นต้น

ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก

ในบรรดาวิธีการแก้ไขจุดอ่อนของข้อสอบเลือกตอบชนิดแบบฉบับที่นักวัดผลเสนอมานี้

แต่ละวิธีก็มีจุดเด่นจุดด้อยแตกต่างกันไป วิธีการที่เห็นว่ามีคุณค่ายิ่งที่จะช่วยแก้ปัญหาจุดอ่อนของข้อสอบแบบเลือกตอบชนิดแบบฉบับได้มากที่สุด น่าจะได้รับการพิจารณานำมาพัฒนาเพื่อการวัดผลการเรียนในชั้นเรียน เพราะมีข้อได้เปรียบหลายประการ ก็คือการใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก (multiple true-false, MTF) ข้อสอบเลือกตอบชนิดนี้ต่างจากข้อสอบเลือกตอบชนิดแบบฉบับและชนิดอื่น ๆ ตรงที่ตัวเลือกที่ให้มาอาจเป็นคำตอบที่ถูกต้องได้มากกว่า 1 ตัวเลือก หรืออาจไม่มีตัวเลือกใดเป็นคำตอบที่ถูกต้องเลย หรืออาจจะถูกต้องทั้งหมดเลยก็ได้ ผู้ตอบจะไม่ทราบมาก่อนว่าแต่ละข้อมีคำตอบที่ถูกต้องอยู่ที่ตัวเลือก ในการตอบผู้ตอบจะต้องตอบให้ได้ว่าตัวเลือกใดบ้างเป็นตัวเลือกที่ถูกต้องและตัวเลือกใดบ้างเป็นตัวเลือกที่ผิด นั่นคือจะต้องตัดสินคำตอบทุกตัวเลือก การให้ตอบในลักษณะนี้เป็นการวัดที่คล้ายกับการประเมินผลการเรียนรู้ในทัศนซึ่งผู้เรียนจะต้องสามารถแยกแยะสิ่งที่เป็นสมาชิกของมโนทัศน์นั้นกับสิ่งที่ไม่ใช่สมาชิกของมโนทัศน์ออกจากกันให้ได้ (Roid, and Haladyna 1982: 150-151) นั่นคือจะต้องรู้ทั้งสิ่งที่ถูกต้องและสิ่งที่ไม่ถูกต้องจึงจะถือว่ารู้จริง นอกจากนี้การให้ตอบข้อสอบในลักษณะนี้เหมือนกับที่เราได้วัดเนื้อหาและพฤติกรรมเดียวกันซ้ำหลาย ๆ ครั้ง โดยไม่ต้องเพิ่มจำนวนข้อสอบและไม่ทำให้ผู้ตอบเกิดความรู้สึกเบื่อหน่ายที่ต้องทำข้อสอบคล้ายกันซ้ำหลายครั้งอีกด้วย ลักษณะเช่นนี้น่าจะทำให้แน่ใจได้ว่าเราวัดความรู้ความสามารถของเขาได้ถูกต้องแม่นยำมากขึ้น และยังสามารถวัดได้ว่าผู้ตอบมีความรู้ความสามารถที่ถูกต้องสมบูรณ์ในเรื่องนั้นหรือไม่ สามารถระบุความแตกต่างของระดับความรู้บางส่วนของผู้เข้าสอบได้ น่าจะเชื่อได้ว่าการใช้ข้อสอบชนิดนี้ช่วยให้การวัดมีความละเอียดถี่ถ้วนมากขึ้น เหมาะที่จะใช้เป็นข้อสอบที่ใช้ตัดสินผลการเรียนและวินิจฉัยข้อบกพร่องในการเรียนได้เป็นอย่างดี ในแง่ของการปฏิบัติก็จะช่วยให้การเขียนข้อสอบง่ายขึ้น เพราะไม่ต้องพะวงกับการพยายามทำตัวเลือกให้มีคำตอบที่ถูกต้องเพียงคำตอบเดียว ซึ่งบางคำถามหรือบางลักษณะวิชาทำได้ยาก จนบางครั้งกลายเป็นการแนะคำตอบช่วยให้ผู้ตอบหาคำตอบได้ง่ายขึ้นเสียอีก ทำให้ผลที่ได้จากการวัดมีส่วนของความคลาดเคลื่อนเสียมากกว่าที่จะเป็นผลจากความสามารถที่แท้จริง ส่วนในเรื่องของการตรวจให้คะแนนก็สามารถทำได้โดยสะดวก เพราะคล้ายกับการตรวจให้คะแนนข้อสอบถูกผิดตามธรรมดา การที่จะนำข้อสอบชนิดนี้มาใช้วัดผลการเรียนให้แพร่หลายจึงน่าจะมีความเป็นไปได้อย่างมาก

ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกนี้ ได้มีการพัฒนาขึ้นและใช้กันมานานแล้วอย่างน้อยก็ก่อนปี ค.ศ. 1939 เพราะในบทความของ Cronbach ในปี 1939 ได้กล่าวถึงข้อสอบชนิดนี้ว่ามีการใช้กันมาแล้ว แต่เรียกชื่อต่าง ๆ กัน เช่น plural multiple-answer

หรือ plural choices เป็นต้น (Cronbach 1939: 628) ต่อมาในปี ค.ศ. 1941 Cronbach ก็ได้ตีพิมพ์ผลการทดลองเปรียบเทียบการใช้ข้อสอบชนิดนี้กับข้อสอบเลือกตอบชนิดหลายคำตอบที่ให้ตอบรับเฉพาะคำตอบที่ถูกเท่านั้น (Cronbach เรียกข้อสอบแบบนี้ว่า multiple multiple-choice ซึ่งที่จริงก็คือข้อสอบที่เรียกกันว่าข้อสอบ multiple-answer ในปัจจุบันนี้ นั่นเอง) การศึกษาของ Cronbach ไม่พบความแตกต่างอย่างมีนัยสำคัญของการตอบข้อสอบ 2 ชนิดนี้ในเรื่องของเวลาที่ใช้ในการตอบ ความยากง่าย ความเที่ยงและความตรงของแบบสอบ Cronbach สรุปว่าในกรณีที่ไม่ต้องการให้ผู้ตอบเว้นไม่ตอบข้อที่ไม่แน่ใจ (อนุญาตให้เดา) แล้ว ข้อสอบทั้ง 2 แบบสามารถใช้ทดแทนกันได้ แต่หลักฐานเชิงประจักษ์ชี้ชวนให้เลือกใช้ข้อสอบแบบหลายคำตอบมากกว่า (Cronbach 1941: 542) อย่างไรก็ตามข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกนี้ไม่ได้รับความสนใจมากนักในวงการวัดผลการศึกษา แต่กลับมีการนำไปใช้กันอย่างแพร่หลายในการทดสอบในวงการแพทยศึกษาในชื่อที่เรียกว่า indeterminate type และเรียกชื่อว่าข้อสอบแบบ multi-facet ในการสอบระดับประโยคมัธยม (Wood 1977: 23)

ในช่วงทศวรรษที่ผ่านมาได้มีรายงานการศึกษาเกี่ยวกับข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกจากวงการแพทยศึกษาและสาขาวิชาที่เกี่ยวข้องกับวิทยาศาสตร์สุขภาพออกมาบ้างแต่ก็ไม่น้อยนัก ทั้งนี้เพื่อที่จะขยายหารูปแบบของข้อสอบแบบใหม่นอกเหนือไปจากข้อสอบเลือกตอบชนิดเชิงซ้อน (complex multiple-choice) ซึ่งใช้กันอยู่ค่อนข้างแพร่หลายในวงการดังกล่าวและพบว่ายังมีประสิทธิภาพในการวัดอยู่ในขั้นที่ไม่น่าพอใจ จากการทบทวนรายงานการศึกษาเรื่องเหล่านี้ของ Frisbie และ Sweeney (1982) พบว่าการรายงานถึงการเหนือกว่าในเรื่องความเที่ยงของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกจากผลงานหลายเรื่อง แต่ Frisbie และ Sweeney เห็นว่าการเปรียบเทียบคุณภาพของแบบสอบในเรื่องขอความเที่ยงของรายงานเหล่านี้ยังไม่เหมาะสม เนื่องจากไม่มีการปรับความยาวของแบบสอบให้เหมาะสมกับเวลาที่ใช้ในการตอบข้อสอบแต่ละชนิด เขาจึงได้ศึกษาเปรียบเทียบคุณภาพของเลือกตอบชนิดแบบฉบับกับข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก ในเรื่องของเวลาที่ใช้ในการตอบข้อสอบ ความเที่ยง ความตรงตามสภาพ ความยากง่าย และความพึงพอใจของผู้ตอบ โดยให้ข้อสอบทั้งสองชนิดมีความเท่าเทียมกันในเรื่องเนื้อหา (content-parallel) พบว่า ในเวลาที่เท่ากันผู้ตอบจะตอบข้อสอบชนิดตัดสินคำตอบทุกตัวเลือกได้ 0.7 เท่าของข้อสอบเลือกตอบชนิดแบบฉบับ เมื่อมีการปรับความยาวของข้อสอบให้เหมาะสมกับเวลาแล้วปรากฏว่า ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกมีความเที่ยงสูงกว่าข้อสอบเลือกตอบชนิดแบบฉบับ สำหรับความตรงนั้น Frisbie และ

Sweeney ได้หาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบทั้งสอง พบว่ามีสหสัมพันธ์กันในระดับค่อนข้างสูง ส่วนเรื่องของความยากง่ายนั้น Frisbie และ Sweeney รายงานว่าไม่อาจสรุปได้ว่าข้อสอบชนิดใดยากง่ายกว่ากัน เนื่องจากข้อสอบที่ใช้ศึกษาครั้งนี้ค่อนข้างยาก ส่วนเรื่องความพึงพอใจของผู้ตอบนั้น ปรากฏว่าผู้ตอบส่วนใหญ่ชอบข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก จากหลักฐานเชิงประจักษ์เหล่านี้ Frisbie และ Sweeney สนับสนุนให้ใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก

การศึกษาของ Frisbie และ Sweeney ดังกล่าวค่อนข้างจะครอบคลุมประเด็นที่จะต้องพิจารณาในการศึกษาคุณภาพของข้อสอบ ยกเว้นเรื่องของอำนาจจำแนกเท่านั้นที่ไม่มีปรากฏในรายงาน ข้อที่น่าสังเกตก็คือทั้งสองพิจารณาข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก ไปใกล้ลักษณะของข้อสอบถูกผิดแบบธรรมดา มากกว่าที่จะมองว่าเป็นข้อสอบแบบเลือกตอบ แล้ววิเคราะห์ข้อมูลจากตัวเลือกต่าง ๆ เหมือนกับเป็นข้อสอบ 1 ข้อ จึงเห็นว่าข้อสอบชนิดตัดสินคำตอบทุกตัวเลือกมีจำนวนข้อมากกว่าถามเนื้อหาสาระที่ต้องการได้ครอบคลุมมากกว่า แต่ความจริงแล้วกลุ่มตัวเลือกกลุ่มหนึ่ง ๆ นั้นวัดเนื้อหาเดียวกันควรเป็นข้อสอบ 1 ข้อ ส่วนการหาความตรงโดยการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบทั้งสองนั้น ไม่น่าจะเป็นดัชนีบอกความตรงของแบบสอบทั้งสองได้ สิ่งที่ได้ควรเป็นข้อสรุปว่าข้อสอบทั้งสองชนิดวัดในสิ่งเดียวกัน ซึ่งก็น่าจะเป็นเช่นนั้นเพราะข้อสอบได้ถูกสร้างให้มีความเท่าเทียมกันในเรื่องเนื้อหาอยู่แล้ว การหาความตรงน่าจะเทียบกันเกณฑ์จากแบบสอบอื่นๆ มากกว่า

ในปี พ.ศ. 2526 จินดา โตอนันต์ (2526) ก็ได้ทำการศึกษาในลักษณะเดียวกับ Frisbie และ Sweeney โดยใช้พลวิชัยที่เป็นนักเรีชนไทย ข้อแตกต่างของงานวิจัยทั้งสองนี้ก็คือจินดาพิจารณาข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก เป็นกลุ่มตัวเลือกคือข้อสอบ 1 ข้อ เทียบกับข้อสอบเลือกตอบชนิดแบบจับ 1 ข้อ จินดาพบว่า ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกง่ายกว่าข้อสอบเลือกตอบชนิดแบบจับและม้ออำนาจจำแนกต่ำกว่า ส่วนเรื่องของความเที่ยงนั้นไม่อาจสรุปได้ว่าข้อสอบชนิดใดมีความเที่ยงสูงกว่า เพราะเนื้อหาส่วนแรกพบว่าข้อสอบชนิดแบบจับมีความเที่ยงสูงกว่า แต่เนื้อหาส่วนที่ 2 กลับพบว่าข้อสอบชนิดตัดสินคำตอบทุกตัวเลือกมีความเที่ยงสูงกว่า ส่วนค่าสหสัมพันธ์ระหว่างแบบสอบทั้งสองชนิดอยู่ในระดับปานกลาง และนักเรียนที่มีความสามารถระดับต่ำชอบและต้องการให้ครูใช้ข้อสอบแบบตัดสินคำตอบทุกตัวเลือกมากกว่านักเรียนที่มีความสามารถระดับปานกลางและสูง

Rosemarie K. Kolstad เป็นอีกผู้หนึ่งที่ทำให้ความสนใจกับข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกนี้มาก ในปี ค.ศ.1982 Kolstad และคณะ (Kolstad, Goaz, and Kolstad 1982) ได้รายงานถึงการใช้ข้อสอบชนิดนี้ (เรียกชื่อว่า nonrestricted multiple-choice) ใน Baylor College of Dentistry ว่าผู้สอนที่ได้ใช้ข้อสอบชนิดนี้รู้สึกชอบข้อสอบข้อสอบชนิดนี้ เพราะเขียนข้อสอบได้ง่ายขึ้นและช่วยลดสิ่งแฉะที่จะเป็นประโยชน์ต่อการเดาในข้อสอบได้ ส่วนผู้สอนที่ไม่เลือกใช้ข้อสอบชนิดนี้ให้เหตุผลว่าข้อสอบชนิดนี้ดูจะยากสำหรับนักศึกษา ในบทความเดียวกันนี้ยังได้กล่าวถึงความเที่ยงของแบบสอบที่ประกอบด้วยข้อสอบชนิดนี้กับของแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดแบบฉบับว่ามีค่าสัมประสิทธิ์ความเที่ยงเป็น 0.72 และ 0.88 ตามลำดับ แต่ค่าความเที่ยงของแบบสอบทั้งสองไม่ควรนำมาเปรียบเทียบกัน เพราะข้อสอบไม่ได้มีความเทียบเท่ากันและเนื้อหาที่แตกต่างกันด้วย อย่างไรก็ตาม Kolstad และคณะก็ได้เสนอว่า ถ้าต้องการประเมินผลการเรียนรู้เชิงสะสม (commulative learning) แล้วน่าจะใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก ต่อมา Kolstad และคณะได้เสนอผลการศึกษาเปรียบเทียบการใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกกับข้อสอบเลือกตอบชนิดเชิงซ้อนถึง 2 เรื่องด้วยกัน (Kolstad, et al. 1983 a,b) ซึ่งให้ผลตรงกันว่า คะแนนเฉลี่ยของกลุ่มผู้ตอบข้อสอบเลือกตอบชนิดเชิงซ้อน สูงกว่าคะแนนเฉลี่ยของผู้ตอบข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก (ซึ่งตรวจให้คะแนนแบบต้องตอบรับหรือปฏิเสธ ได้ถูกต้องทุกตัวเลือก ในข้อนี้จึงจะได้คะแนน 1 คะแนน) Kolstad และคณะสรุปว่า ข้อสอบเลือกตอบชนิดเชิงซ้อนวัดความสามารถ ได้ค่าที่สูงกว่าที่ควรจะเป็น เพราะมีสิ่งแฉะให้ผู้ตอบที่มีความรู้บางส่วนเดาคำตอบได้สำเร็จ ในการสอบวัดควรจะใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกมากกว่า แต่จากการศึกษาของ Kolstad และคณะทั้ง 2 เรื่องใช้ข้อสอบจำนวนค่อนข้างน้อย (14 และ 10 ข้อ) และศึกษาเปรียบเทียบในแง่ของคะแนนเฉลี่ยเพียงอย่างเดียว จึงไม่อาจสรุปผลเกี่ยวกับคุณภาพของข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก ได้มากนัก

จากการติดตามรายงานการศึกษาวิจัยเกี่ยวกับข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกพอสรุปได้ว่า การศึกษาวิจัยเกี่ยวกับข้อสอบชนิดนี้ยังมีค่อนข้างน้อยเมื่อเทียบกับข้อสอบชนิดอื่น ๆ ผลจากการวิจัยก็ยังไม่ค่อยจะสอดคล้องกัน เช่น เรื่องของความเที่ยง Frisbie และ Sweeney พบว่า ข้อสอบชนิดนี้มีความเที่ยงค่อนข้างสูง (ประมาณ 0.75 ขึ้นไป) และสูงกว่าข้อสอบเลือกตอบชนิดแบบฉบับ แต่ข้อค้นพบของจินตากลันพบว่า ความเที่ยงอยู่ในระดับปานกลาง (0.42 หรือ 0.64) และไม่อาจสรุปได้ว่ามีความเที่ยงสูงหรือต่ำกว่าข้อสอบเลือกตอบชนิดแบบฉบับ

ส่วน Cronbach นั้นพบว่าความเที่ยงอยู่ในระดับปานกลางค่อนข้างต่ำ (.428) เรื่องของความยากง่าย Frisbie และ Sweeney ไม่อาจสรุปได้ว่าข้อสอบชนิดใดยากง่ายกว่ากัน แต่จินดาสรุปว่า ข้อสอบเลือกตอบชนิดตัดลีนคำตอบทุกตัวเลือกว่าง่ายกว่าข้อสอบเลือกตอบชนิดแบบจับ ส่วน Cronbach พบว่าความยากง่ายไม่ต่างจากข้อสอบเลือกตอบชนิดเลือกหลายคำตอบ สำหรับค่าอำนาจจำแนกนั้นจินดารายงานว่า มีค่าอำนาจจำแนกค่อนข้างต่ำและต่ำกว่าข้อสอบเลือกตอบชนิดแบบจับ ส่วนคนอื่น ๆ ไม่มีรายงานเรื่องค่าอำนาจจำแนก ในเรื่องของความตรง Cronbach รายงานว่ามีความตรงพอประมาณ (.598) ส่วนการศึกษาของ Frisbie และ Sweeney นั้นไม่น่าสรุปว่าเป็นความตรงของแบบสอบได้ จากข้อค้นพบเหล่านี้ยังไม่อาจสรุปข้อความรู้เกี่ยวกับคุณภาพหรือประสิทธิภาพของข้อสอบเลือกตอบชนิดตัดลีนคำตอบทุกตัวเลือกได้ชัดเจน โดยเฉพาะประสิทธิภาพเมื่อเทียบกับข้อสอบเลือกตอบชนิดแบบจับ

ทฤษฎีการวัดที่นำมาใช้ศึกษาประสิทธิภาพของข้อสอบ

เมื่อพิจารณาถึงกรอบทฤษฎีการวัดที่ใช้เป็นแนวทางในการศึกษาเกี่ยวกับประสิทธิภาพของข้อสอบเลือกตอบชนิดตัดลีนคำตอบทุกตัวเลือกหรือข้อสอบแบบอื่น ๆ ที่ศึกษากันมา เกือบทั้งหมดอาศัยทฤษฎีการวัดแบบคลาสสิก (classical test theory) การพิจารณาเปรียบเทียบคุณภาพข้อสอบจะแยกพิจารณาเป็นประเด็น ๆ ไปอย่างอิสระจากกัน (เช่น เรื่องความตรง ความเที่ยง ความยากง่าย อำนาจจำแนก ฯลฯ) บางครั้งพบว่าคุณภาพในประเด็นหนึ่งดีกว่า แต่กลับน้อยกว่าในอีกประเด็นหนึ่ง ทำให้ยากแก่การสรุปข้อความรู้เกี่ยวกับคุณภาพของข้อสอบชนิดนั้น ๆ ได้นอกจากนั้นค่าสถิติต่าง ๆ ของข้อสอบที่วิเคราะห์โดยอาศัยกรอบทฤษฎีการวัดแบบคลาสสิกมักจะมีค่าแปรเปลี่ยนไปตามกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกัน ซึ่งเป็นจุดอ่อนที่ทำให้การวิเคราะห์ข้อสอบแบบคลาสสิกไม่เป็นที่ยอมรับจากนักวัดผลในยุคปัจจุบัน (Lord 1980: 35) ทฤษฎีการวัดแบบคลาสสิกเองก็ได้รับการวิจารณ์ว่าเป็นทฤษฎีที่ตั้งอยู่บนข้อตกลงเบื้องต้นที่อ่อนและไม่ค่อยจะสมเหตุสมผล (Lord 1980: 9; Allen and Yen 1979: 57 - 60) นอกจากนี้ยังมีจุดอ่อนอื่น ๆ อีก เช่น การเปรียบเทียบความสามารถระหว่างบุคคลทำได้ในวงจำกัดเฉพาะเป็นผลการวัดด้วยข้อสอบชุดเดียวกันหรือคู่ขนานกันเท่านั้น มโนทัศน์ของแบบสอบคู่ขนานเป็นสิ่งที่เป็นไปได้ยากในทางปฏิบัติ ไม่มีสารสนเทศที่เป็นพื้นฐานที่จะช่วยทำนายว่าพฤติกรรมการตอบสนองของผู้ตอบจะเป็นอย่างไรเมื่อเขาต้องตอบข้อสอบข้อหนึ่ง ๆ และมีการสมมุติว่าความแปรปรวนของความคลาดเคลื่อนในการวัดมีค่าเท่ากันสำหรับผู้ตอบทุกคน ซึ่งไม่สอดคล้องกับสภาพความเป็นจริง

(Hambleton, and Swaminathan 1985: 2-3) จึงทำให้ข้อความรู้ที่ได้จากการวิจัยโดยอาศัยกรอบทฤษฎีนี้เป็นแนววิเคราะห์เป็นที่น่าสงสัย นำได้รับการตรวจสอบด้วยวิธีการวิเคราะห์ที่อาศัยกรอบทฤษฎีใหม่ที่อธิบายปรากฏการณ์เกี่ยวกับการวัดได้ดีกว่า

ทฤษฎีการตอบสนองต่อข้อสอบ (item response theory) เป็นทฤษฎีการวัดทฤษฎีหนึ่งที่เริ่มเข้ามามีบทบาทในวงการวัดทางการศึกษาและจิตวิทยาตั้งแต่ทศวรรษ 1950-1960 และมีบทบาทเพิ่มมากขึ้นเรื่อย ๆ ในปัจจุบัน เนื่องจากทฤษฎีนี้สามารถแก้จุดอ่อนของทฤษฎีการวัดแบบคลาสสิกได้และสามารถนำไปประยุกต์เพื่อแก้ปัญหาการวัดทางการศึกษาและจิตวิทยาได้หลายเรื่อง (Lord 1977: 117-138) ซึ่งทฤษฎีการวัดแบบคลาสสิกไม่สามารถจะแก้ปัญหาได้ ทฤษฎีนี้มีความเชื่อว่าพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบนั้นถูกกำหนดโดยลักษณะบางอย่างภายในตัวบุคคล (latent trait) ซึ่งไม่สามารถจะสังเกตได้ จึงพยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคล (trait) กับพฤติกรรมการตอบสนองต่อข้อสอบ (performance) การอธิบายความสัมพันธ์ดังกล่าวใช้รูปของแบบจำลองทางคณิตศาสตร์ได้หลายรูปแบบ จุดประสงค์ที่สำคัญของทฤษฎีนี้คือมุ่งที่จะสรุปเกี่ยวกับลักษณะภายในตัวของบุคคลโดยอ้างอิงจากพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น โดยอาศัยแบบจำลองทางคณิตศาสตร์เป็นตัวเชื่อมที่สำคัญ จุดเด่นของทฤษฎีนี้ก็คือ ค่าพารามิเตอร์ของข้อสอบไม่ขึ้นอยู่กับลักษณะของกลุ่มตัวอย่างผู้ตอบข้อสอบ การประมาณค่าความสามารถของบุคคลไม่ขึ้นอยู่กับกลุ่มตัวอย่างของข้อสอบ นั่นคือแม้ผู้สอบจะตอบข้อสอบที่ไม่ใช่ชุดเดียวกันก็สามารถประมาณค่าความสามารถเปรียบเทียบกันได้ และที่สำคัญก็คือสามารถหาความแม่นยำของการประมาณค่าความสามารถของบุคคลได้ (Hambleton, and Swaminathan 1985: 11)

คำสารสนเทศ : ดัชนีชี้ความแม่นยำของการประมาณค่าความสามารถ

นอกจากจะให้แนวทางในการแก้ปัญหาเกี่ยวกับการวัดในหลาย ๆ เรื่องแล้ว ทฤษฎีการตอบสนองต่อข้อสอบนี้ยังได้ให้แนวคิดที่ชัดเจนเกี่ยวกับการเปรียบเทียบประสิทธิภาพของแบบสอบไว้ด้วย การเปรียบเทียบประสิทธิภาพของแบบสอบตามทฤษฎีนี้มาจากแนวความคิดที่ว่า ผู้วัดใช้แบบสอบวัดความสามารถของบุคคลโดยให้บุคคลตอบสนองต่อข้อสอบแล้ว ใช้ผลการตอบสนองนั้นสรุปอ้างอิงเกี่ยวกับความสามารถที่มีอยู่ในตัวบุคคล ดังนั้นแบบสอบที่ดีหรือมีคุณภาพจึงควรเป็นแบบสอบที่ให้ผลการวัดซึ่งสามารถนำไปใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำมากที่สุด

ดัชนีที่ใช้วัดระดับความถูกต้องแม่นยำ (precision) ในการประมาณค่าความสามารถที่ต้องการวัด ก็คือค่าสารสนเทศ (information function) (Birnbaum 1968: 419) โดยที่ค่าสารสนเทศนี้จะมีค่ามากหรือน้อยขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบซึ่งเป็นตัวกำหนดลักษณะของข้อสอบ และค่าสารสนเทศนี้จะมีค่าแตกต่างกันถ้าเราใช้ผลจากการวัดด้วยข้อสอบข้อเดียวกันหรือแบบสอบฉบับเดียวกันประมาณค่าความสามารถในระดับที่แตกต่างกัน ณ ระดับความสามารถใดที่ค่าสารสนเทศที่ได้จากข้อสอบหรือแบบสอบนั้นมีค่าสูง แสดงว่าช่วงของค่าประมาณจะแคบ นั่นหมายความว่า ความถูกต้องแม่นยำในการประมาณค่าความสามารถที่ระดับนั้นจะมีมาก ในทางกลับกันถ้าค่าสารสนเทศ ณ ระดับความสามารถใดมีค่าน้อยก็แสดงว่า ช่วงของการประมาณค่าจะกว้าง ซึ่งก็หมายความว่า ความถูกต้องแม่นยำในการประมาณค่าความสามารถระดับนั้นก็น้อย ดังนั้นถ้าเรามีแบบสอบหลายฉบับที่วัดความสามารถอย่างเดียวกันแล้ว เราก็จะสามารถเปรียบเทียบคุณภาพหรือประสิทธิภาพของแบบสอบเหล่านั้นได้ โดยการนำเอาค่าสารสนเทศ ณ ระดับความสามารถเดียวกันของแบบสอบที่ต้องการเปรียบเทียบกันมาหาค่าอัตราส่วนกัน ค่าอัตราส่วนระหว่างค่าสารสนเทศของแบบสอบ 2 ฉบับใด ๆ ที่ระดับความสามารถใด ก็คือ ค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์ (relative efficiency) ของแบบสอบฉบับหนึ่งเมื่อเทียบกับแบบสอบอีกฉบับหนึ่ง ณ ระดับความสามารถนั้น ๆ ถ้าค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์มีค่าเท่ากับ 1 แสดงว่าแบบสอบทั้งสองมีประสิทธิภาพในการประมาณค่าความสามารถเท่ากัน ถ้าค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์มีค่าน้อยกว่าหรือมากกว่า 1 ก็แสดงว่าแบบสอบทั้งสองมีประสิทธิภาพในการประมาณค่าความสามารถไม่เท่ากัน ส่วนแบบสอบใดจะมีประสิทธิภาพมากกว่านั้นขึ้นอยู่กับว่าเราคำนวณค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์ของแบบสอบใดเทียบกับแบบสอบใด วิธีการนี้สามารถใช้เปรียบเทียบประสิทธิภาพของข้อสอบในระดับรายข้อได้ด้วย (Hambleton, and Swaminathan 1985: 123) วิธีการเปรียบเทียบเชิงสัมพัทธ์โดยการใช้ค่าอัตราส่วนนี้ เป็นวิธีที่นิยมใช้กันมากในการเปรียบเทียบประสิทธิภาพของสิ่งต่าง ๆ เช่น การเปรียบเทียบประสิทธิภาพของค่าสถิติทดสอบ (test statistic) ในสถิติแบบนอนพารามิเตอร์ (non-parametric statistics) (Marascuilo, and McSweeney 1977: 77-87) การเปรียบเทียบประสิทธิภาพของตัวประมาณค่า (estimator) พารามิเตอร์ต่าง ๆ หรือการหาประสิทธิภาพของเครื่องกลต่าง ๆ ในวิชากลศาสตร์ เป็นต้น จะเห็นว่าการเปรียบเทียบคุณภาพของแบบสอบโดยใช้ค่าสารสนเทศเป็นตัวกำหนดคุณภาพของแบบสอบเป็นการเปรียบเทียบคุณภาพของแบบสอบโดยสรุปจากทุกแง่มุมเนื่องจากค่าสารสนเทศเป็นฟังก์ชันของค่าพารามิเตอร์ของข้อสอบ จึงมีลักษณะเป็นดัชนีผสม (composite index) ที่สร้างจากดัชนีที่กำหนดคุณภาพของข้อสอบหลาย ๆ ดัชนีด้วยกัน การ

เปรียบเทียบคุณภาพของข้อสอบหรือแบบสอบจึง ไม่ต้องเปรียบเทียบคุณลักษณะที่ละประเด็นอีกต่อไป

นอกจากจะใช้ผลการวัดด้วยแบบสอบประมาณค่าความสามารถที่มีอยู่ในตัวบุคคลว่าเรามีความสามารถอยู่ในระดับใดแล้ว ในสถานการณ์ที่เกี่ยวกับการเรียนการสอนเรายังใช้คะแนนจากแบบสอบช่วยตัดสินผลการเรียนรู้ของผู้เรียนว่ามีความรู้ความสามารถถึงระดับที่ต้องการหรือไม่การใช้แบบสอบช่วยตัดสินใจในลักษณะเช่นนี้เป็นการใช้แบบสอบในลักษณะเป็นแบบสอบวัดการเรียนรู้เพื่อรู้จริง (Mastery Learning) การเรียนเพื่อรู้จริงนี้เป็นระบบที่พึงปรารถนาในการจัดการเรียนการสอน ซึ่งมีผู้เห็นความสำคัญและนำมาใช้กันมากขึ้นในวงการศึกษาในประเทศไทย การใช้แบบสอบในลักษณะเป็นแบบสอบวัดการเรียนรู้เพื่อรู้จริงจึงมีบทบาทและความสำคัญมากขึ้นเรื่อย ๆ การศึกษาประสิทธิภาพของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกเมื่อเทียบกับข้อสอบเลือกตอบชนิดแบบฉบับนอกจากจำศึกษาประสิทธิภาพในการประมาณค่าความสามารถที่แท้จริงแล้ว น่าจะได้ศึกษาถึงประสิทธิภาพของแบบสอบที่ประกอบด้วยข้อสอบทั้งสองชนิด เมื่อใช้เป็นแบบสอบเพื่อวัดการเรียนรู้เพื่อรู้จริงประกอบด้วย

ในการศึกษาประสิทธิภาพของแบบสอบที่ใช้วัดการเรียนรู้เพื่อรู้จริง (mastery test) นั้นสามารถศึกษาได้หลายวิธี วิธีที่น่าสนใจในขณะนี้คือ วิธีการศึกษาตามแนวคิดของ Huyhn (1980: 385-406) ซึ่งเป็นแนวคิดที่อาศัยกรอบทฤษฎีการตอบสนองต่อข้อสอบผสมกับทฤษฎีการตัดสินใจทางสถิติ (statistical decision theory) โดยจะเริ่มจากนำข้อสอบที่ทราบค่าพารามิเตอร์แล้วมาคำนวณหาคะแนนจุดตัดด้วยเกณฑ์มินิมแม็กซ์ (minimax) แล้วเปรียบเทียบค่าฟังก์ชันการเสี่ยง (risk function) ต่อการตัดสินผิดพลาด (คือตัดสินให้ผู้ที่รู้จริง (master) เป็นผู้สอบไม่ผ่าน หรือตัดสินให้ผู้ที่ไม่รู้จริง (nonmaster) เป็นผู้สอบผ่าน) โดยใช้คะแนนจากแบบสอบ กับค่าฟังก์ชันการเสี่ยงต่อการตัดสินผิดพลาด โดยที่ไม่ใช้คะแนนจากแบบสอบเลย แบบสอบจะมีประสิทธิภาพในการใช้เป็นแบบสอบวัดการเรียนรู้เพื่อรู้จริงหรือไม่อยู่ที่ว่า การตัดสินผลการเรียนโดยใช้คะแนนจากแบบสอบช่วยลดความเสี่ยงต่อการตัดสินผิดพลาดที่ตัดสินโดยที่ไม่ใช้คะแนนจากแบบสอบได้มากน้อยเพียงใด อัตราส่วนของความเสี่ยงที่ลดลงต่อความเสี่ยงต่อการตัดสินผิดพลาดเมื่อไม่ได้ใช้แบบสอบ จะเป็นดัชนีบอกประสิทธิภาพของการตัดสินผลการเรียนเพื่อรู้จริง (decision efficiency)

อนึ่งข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกอาจจะมีวิธีการตรวจให้คะแนนได้หลายวิธี ถ้าจะแบ่งแนวคิดของการตรวจให้คะแนนข้อสอบชนิดนี้ก็จะแบ่งได้ 3 แนวคือ กลุ่มแรกจะยอมรับเฉพาะผู้ที่มีความรู้ถูกต้องสมบูรณ์เท่านั้น การให้คะแนนตามแนวคิดนี้ผู้ตอบจะต้องตอบรับและตอบปฏิเสธตัวเลือกทุกตัวได้ถูกต้องทุกตัวเลือกจึงจะได้คะแนน 1 คะแนนในข้อนั้น ถ้าตอบรับหรือปฏิเสธตัวเลือกใดตัวเลือกหนึ่งผิดอย่างน้อย 1 ตัวเลือกก็จะไม่ได้คะแนนเลย การให้คะแนนตามแนวคิดแรกนี้ถือเป็นการให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์เท่านั้น กลุ่มที่ 2 เป็นพวกที่ยอมรับผู้ที่มีความรู้บางส่วน การให้คะแนนตามแนวคิดนี้จะให้คะแนนแก่ทุกตัวเลือกที่ตอบรับหรือปฏิเสธได้ถูกต้อง ตัวเลือกละเท่า ๆ กัน แต่ถ้าตอบผิดหรือเว้นไว้ก็จะไม่ได้คะแนน กลุ่มที่ 3 ให้มีการลงโทษเมื่อมีการตอบรับหรือปฏิเสธ ไม่ถูกต้อง จะต้องมีการกำหนดสูตรในการให้คะแนนซึ่งอาจทำได้หลายแบบ สำหรับการวิจัยนี้ผู้วิจัยสนใจวิธีการตรวจให้คะแนนตามแนวคิดของสองกลุ่มแรก เพราะเป็นวิธีการที่ไม่ยุ่งยากสะดวกที่จะนำมาใช้ในการวัดผลการเรียนในชั้นเรียน จึงมุ่งที่จะเปรียบเทียบประสิทธิภาพของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกเมื่อมีการตรวจให้คะแนนทั้ง 2 แบบนี้ด้วย

ด้วยรูปแบบของข้อสอบและวิธีการตอบเมื่อพิจารณาในเชิงเหตุผลแล้ว จะเห็นว่าข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกน่าจะแก้ปัญหาจุดอ่อนของการใช้ข้อสอบเลือกตอบชนิดแบบฉบับและน่าจะเป็นข้อสอบที่มีประสิทธิภาพสูงในการวัดผลการเรียน เพราะมีคำตอบที่ถูกได้หลายคำตอบผู้ตอบจะต้องจำแนกให้ได้ว่าตัวเลือกใดผิดและตัวเลือกใดถูก การตรวจให้คะแนนสำหรับผู้ที่มีความรู้บางส่วนก็สามารถทำได้ ทำให้การวัดมีความละเอียดถี่ถ้วนมากขึ้น สามารถแยกผู้มีความรู้ในระดับต่าง ๆ กันออกจากกันได้ละเอียดขึ้น น่าจะทำให้การวัดมีความถูกต้องแม่นยำมากขึ้น อีกทั้งการนำไปใช้ในสถานการณ์จริงมีความเป็นไปได้อย่างมาก เพราะช่วยให้เขียนข้อสอบได้ง่ายขึ้น การตอบ การตรวจให้คะแนนก็ไม่ได้มีความยุ่งยากไปกว่าการใช้ข้อสอบเลือกตอบชนิดแบบฉบับเลย จึงเหมาะที่นำมาใช้วัดผลการเรียนในชั้นเรียน แต่หลักฐานเชิงประจักษ์เกี่ยวกับประสิทธิภาพของข้อสอบชนิดนี้จากรายงานการศึกษาวิจัยต่าง ๆ ที่มีอยู่ ยังไม่สามารถสรุปเป็นข้อความรู้ที่สมบูรณ์และชัดเจนได้ จึงน่าจะได้มีการศึกษาถึงประสิทธิภาพของข้อสอบชนิดนี้อีกครั้งหนึ่ง โดยอาศัยกรอบทฤษฎีการตอบสนองต่อข้อสอบมาเป็นแนวทางในการวิเคราะห์เปรียบเทียบ เพื่อจะได้ข้อความรู้เกี่ยวกับประสิทธิภาพของข้อสอบชนิดนี้สมบูรณ์และชัดเจนขึ้น

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์เท่านั้น กับที่ให้คะแนนแก่ความรู้บางส่วน เมื่อเทียบกับข้อสอบเลือกตอบชนิดแบบฉบับ ทั้งในระดับรายชื่อและในระดับแบบสอบ
2. เพื่อศึกษาประสิทธิภาพของการตัดสินผลการเรียนเพื่อรู้อัจฉริยะของแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์เท่านั้นกับที่ให้คะแนนแก่ความรู้บางส่วนและแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดแบบฉบับ เมื่อใช้เป็นแบบสอบวัดการ เรียนเพื่อรู้อัจฉริยะ

สมมติฐานของการวิจัย

เนื่องจากการตอบข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก 1 ข้อ ผู้ตอบจะต้องพิจารณาตัดสินใจตอบรับหรือปฏิเสธตัวเลือกทีละตัวเลือก โดยมีคำถามหรือตัวปัญหาอันเดียวกันทุกตัวเลือก จึงเปรียบเสมือนว่าผู้ตอบต้องตอบข้อสอบข้อเดียวกันซ้ำหลายๆ ครั้ง ส่วนการตอบข้อสอบเลือกตอบชนิดแบบฉบับ 1 ข้อ ก็เหมือนกับเป็นการวัดเรื่องนั้นเพียงครั้งเดียว ตามหลักของการรวมหน่วย (The Principle of Aggregation) ซึ่งกล่าวว่า "ผลรวมของชุดของการวัดหลายๆ ครั้งจะเป็นตัวประมาณค่าที่มีเสถียรภาพและปราศจากอคติมากกว่าผลจากการวัดเพียงครั้งเดียว" (the sum of a set of multiple measurements is a more stable and unbiased estimator than any single measurement from the set) (Rushton, Brainerd, and Pressley 1983: 18-19) นั่นคือการวัดสิ่งเดียวกันซ้ำหลายๆ ครั้งย่อมทำให้ผลการวัดมีความแน่นอนเชื่อถือได้มากขึ้น เมื่อใช้คะแนนที่เป็นผลจากการวัดเช่นนี้ประมาณค่าความสามารถที่แท้จริงของผู้ตอบก็น่าจะได้ค่าประมาณที่มีความถูกต้องแม่นยำกว่าการวัดเพียงครั้งเดียว การให้คะแนนแก่ความรู้บางส่วนก็จะทำให้ผลการวัดมีความละเอียดถี่ถ้วนมากขึ้น ส่วนการให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์นั้นผู้ที่จะได้คะแนนแต่ละข้อจะต้องตัดสินใจตอบรับหรือปฏิเสธให้ถูกต้องทุกตัวเลือก ซึ่งคงจะต้องใช้ความสามารถมากขึ้น ผู้ที่มีความสามารถในระดับสูงเท่านั้นจึงจะมีโอกาสตอบถูกได้มาก โอกาสที่ผู้มีความรู้ระดับปานกลางหรือต่ำจะตอบถูกหรือได้คะแนนน่าจะน้อย ด้วยเหตุผลเช่นนี้ผู้วิจัยจึงตั้งสมมติฐานของการวิจัยดังนี้

สมมติฐานที่ 1 ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้บางส่วนจะมีประสิทธิภาพสูงกว่า ข้อสอบเลือกตอบชนิดแบบฉบับในทุกระดับความสามารถ ทั้งในระดับรายชื่อและระดับแบบสอบ

สมมติฐานที่ 2 ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ จะมีประสิทธิภาพสูงกว่า ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้บางส่วนและข้อสอบเลือกตอบชนิดแบบฉบับ เมื่อใช้ประมาณค่าความสามารถของผู้ตอบที่มีความสามารถในระดับสูง ($\theta \geq 2$) แต่จะมีประสิทธิภาพต่ำกว่า เมื่อใช้ประมาณค่าความสามารถของผู้ตอบที่มีความสามารถระดับปานกลางและต่ำ ($\theta < 2$)

สมมติฐานที่ 3 ค่าดัชนีประสิทธิภาพของการตัดสินผลการเรียนเพื่อรูัจริงโดยให้คะแนนจากแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้บางส่วนจะสูงกว่าการใช้คะแนนจากแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ และจากแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดแบบฉบับ

สมมติฐานที่ 4 ค่าดัชนีประสิทธิภาพของการตัดสินผลการเรียนเพื่อรูัจริงโดยให้คะแนนจากแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ จะสูงกว่าการใช้คะแนนจากแบบสอบที่ประกอบด้วยข้อสอบเลือกตอบชนิดแบบฉบับ

ขอบเขตของการวิจัย

ตัวแปรที่ศึกษา ได้แก่

1. รูปแบบของข้อสอบกับวิธีการตรวจให้คะแนน มี 3 ระดับคือ
 - 1.1 ข้อสอบเลือกตอบชนิดแบบฉบับ
 - 1.2 ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก ตรวจให้คะแนนแก่ความรู้บางส่วน
 - 1.3 ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก ตรวจให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์
2. ค่าสารสนเทศของข้อสอบ
3. ค่าสารสนเทศของแบบสอบ

4. ค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบ
5. ค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์ของแบบสอบ
6. ค่าดัชนีประสิทธิภาพของการตัดสินใจของแบบสอบเมื่อใช้เป็นแบบวัดการเรียนรู้ เพื่อรู้จริง

แบบจำลองของทฤษฎีการตอบสนองต่อข้อสอบที่ใช้ในการศึกษาคั้งนี้ ใช้แบบ Three-Parameter Logistic Model

แบบสอบที่ใช้ในการศึกษาคั้งนี้ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาศึกษา 251 การประเมินผลและการสร้างแบบทดสอบ ตามหลักสูตรประกาศนียบัตรวิชาการศึกษาชั้นสูง และหลักสูตรปริญญาครุศาสตร์บัณฑิตของสภาการฝึกหัดครู นุทศศิกราช 2524

กลุ่มประชากรที่ใช้เป็นผู้ตอบข้อสอบ คือ นักศึกษาวิทยาลัยครูในหลักสูตรประกาศนียบัตรวิชาการศึกษาชั้นสูง หรือหลักสูตรปริญญาครุศาสตร์บัณฑิต ที่ลงทะเบียนเรียนวิชาศึกษา 251 การประเมินผลและการสร้างแบบทดสอบ

สาเหตุที่เลือกใช้นี้เนื้อหาวิชา การประเมินผลและการสร้างแบบทดสอบ (ศึกษา 251) มาสร้างแบบสอบที่จะศึกษาและใช้กลุ่มประชากรเป็นนักศึกษาวิทยาลัยครู เนื่องจากคุณภาพของข้อมูลในการวิจัยนี้ขึ้นอยู่กับคำตอบข้อสอบอย่างเต็มความสามารถของผู้ตอบเป็นสำคัญ ซึ่งจะต้องอาศัยความร่วมมือทั้งจากผู้ดำเนินการสอบและผู้ตอบข้อสอบ จึงต้องใช้ผู้ดำเนินการสอบที่สามารถจูงใจให้ผู้ตอบเห็นความสำคัญของการสอบและตอบข้อสอบอย่างเต็มความสามารถ อาจารย์ผู้สอนวิชาศึกษา 251 ในวิทยาลัยครูทั้งหมด เป็นผู้ที่มีความรู้ความสามารถในเรื่องของการวัดและประเมินผลการศึกษา อยู่ในแวดวงวิชาการด้านนี้ มีความสนใจติดตามความรู้ใหม่ ๆ และให้การสนับสนุนเกี่ยวกับงานวิจัยด้านการวัดและประเมินผลค่อนข้างมาก ผู้วิจัยเชื่อว่าเมื่ออาจารย์ผู้สอนเหล่านี้ได้รับการขอร้องให้เป็นผู้ดำเนินการสอบตามวิธีการผู้วิจัยก็จะให้ความร่วมมือช่วยเหลือเป็นอย่างดีและสามารถช่วยกระตุ้นให้นักศึกษาที่เป็นผลวิจัยตั้งใจทำข้อสอบอย่างเต็มความสามารถและดำเนินการสอบได้อย่างมีประสิทธิภาพ ส่วนเนื้อหาวิชาศึกษา 251 เองก็มีลักษณะเป็นโครงสร้างที่ชัดเจนสมควรที่จะออกข้อสอบได้ทั้ง 2 รูปแบบ และเป็นวิชาบังคับของหลักสูตรที่ทุกคนต้องเรียน นอกจากนี้ผู้วิจัยยังเคยเป็นผู้สอนวิชานี้มาก่อน มีความคุ้นเคยและรู้เนื้อหาวิชานี้เป็นอย่างดีซึ่งจะ

ช่วยให้ข้อสอบที่สร้างขึ้นมีความตรงสูง และการเขียนข้อสอบมีคุณภาพได้ในระดับหนึ่ง อีกทั้งการประมาณค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองต่อข้อสอบนั้น ค่าพารามิเตอร์ที่ได้จะไม่ใช่ขึ้นอยู่กับลักษณะของกลุ่มตัวอย่างแต่อย่างใด

ข้อตกลงเบื้องต้น

การสอบในเวลา สถานที่และมีผู้ดำเนินการสอบแตกต่างกันไม่มีผลต่อการตอบสนองต่อข้อสอบของนักศึกษาที่เป็นผลวิจัย

นิยามศัพท์ที่ใช้ในการวิจัย

ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบ	หมายถึง	อัตราส่วนระหว่างความแม่นยำในการประมาณค่าความสามารถจริงด้วยผลการตอบข้อสอบที่นำมาเปรียบเทียบกัน
ประสิทธิภาพเชิงสัมพัทธ์ของแบบสอบ	หมายถึง	อัตราส่วนระหว่างความแม่นยำในการประมาณค่าความสามารถจริงด้วยคะแนนที่เป็นผลจากการตอบแบบสอบที่นำมาเปรียบเทียบกัน
ค่าสารสนเทศของข้อสอบ	หมายถึง	ความแม่นยำในการประมาณค่าความสามารถจริงด้วยผลการตอบข้อสอบข้อนั้น ๆ มีค่าเท่ากับอัตราส่วนของกำลังสองของความชันของเส้นโค้งลักษณะของข้อสอบต่อค่าความแปรปรวนของข้อสอบข้อนั้น

$$I(\theta, u_1) = [P'_1(\theta)]^2 / P_1(\theta)Q_1(\theta)$$

เมื่อ $P_1(\theta)$ คือค่าฟังก์ชันของเส้นโค้งลักษณะของข้อสอบ ณ ระดับ ความสามารถ θ

$$Q_1(\theta) = 1 - P_1(\theta)$$

ค่าสารสนเทศของแบบสอบ	หมายถึง	ความแม่นยำในการประมาณค่าความสามารถจริง ด้วยคะแนนจากการตอบแบบสอบ (Y) ซึ่งมีค่าเท่ากับอัตราส่วนของกำลังสองของความชันของเส้นถดถอยของคะแนน Y บนคะแนนความสามารถจริง (θ) ต่อความแปรปรวนของคะแนน Y ณ ระดับความสามารถ θ
----------------------	---------	---

$$I(\theta, y) = [d\mu_{y|\theta} / d\theta]^2 / \text{Var}(y|\theta)$$

ข้อสอบเลือกตอบชนิดแบบจับ	หมายถึง	ข้อสอบปรนัยที่กำหนดตัวเลือกไว้หลายตัวเลือก แต่จะมีตัวเลือกที่เป็นคำตอบที่ถูกต้องหรือตัวเฉลยเพียง 1 ตัวเลือก ผู้ตอบจะต้องเลือกคำตอบจากตัวเลือกที่กำหนดให้เพียงค่าเดียว ถ้าเลือกได้ตรงกับตัวเฉลยก็จะได้คะแนน 1 คะแนน ถ้าเลือกผิดก็จะได้คะแนน 0 คะแนน ข้อสอบนี้จะสร้างขึ้นตามหลักการเขียนข้อสอบ ผ่านการพิจารณาของผู้เชี่ยวชาญ และมีคุณภาพตามเกณฑ์การวิเคราะห์ข้อสอบแบบสอบคลาสสิก
ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก	หมายถึง	ข้อสอบปรนัยที่กำหนดตัวเลือกไว้ให้หลายตัวเลือกในบรรดาตัวเลือกเหล่านั้นอาจจะมีคำตอบที่ถูกต้องอยู่ได้หลายตัวหรืออาจจะเป็นทั้งหมดหรือไม่มีเลยก็ได้ ผู้ตอบจะต้องตอบรับตัวเลือกที่เป็นคำตอบที่ถูกต้องทุกตัวและตอบปฏิเสธตัวเลือกที่เป็นคำตอบที่ผิดทุกตัว
การตรวจให้คะแนนแก่ความรู้บางส่วน	หมายถึง	การตรวจให้คะแนนข้อสอบตอบชนิดตัดสินคำตอบทุกตัวเลือก แก่ตัวเลือกทุกตัว โดยตัวเลือกใดที่เป็นคำตอบที่ถูกต้องถ้าผู้ตอบตอบรับก็จะได้คะแนน 1 คะแนน ถ้าตอบปฏิเสธก็จะได้ 0 คะแนน ส่วนตัวเลือกที่เป็นคำตอบที่ผิดถ้าผู้ตอบตอบรับก็จะได้

<p>การตรวจให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์</p>	<p>0 คะแนน แต่ถ้าตอบปฏิเสธก็จะได้ 1 คะแนน คะแนนสำหรับข้อคำถามหนึ่ง ก็คือคะแนนรวมจาก ตัวเลือกทุกตัวในข้อนั้น</p> <p>หมายถึง การตรวจให้คะแนนข้อสอบเลือกตอบชนิด ตัดสินคำตอบทุกตัวเลือก โดยผู้ตอบจะต้องตอบรับ สำหรับตัวเลือกที่เป็นคำตอบที่ถูกต้อง และต้อง ปฏิเสธสำหรับตัวเลือกที่เป็นคำตอบที่ผิดได้ทุกตัว เลือก จึงจะได้คะแนน 1 คะแนน แต่ถ้าตอบกลับ กันในตัวเลือกใดตัวเลือกหนึ่งอย่างน้อย 1 ตัว เลือกก็จะได้คะแนน 0 ในข้อคำถามนั้น</p>
<p>ข้อสอบ T.MC</p>	<p>หมายถึง ข้อสอบเลือกตอบชนิดแบบฉบับ</p>
<p>ข้อสอบ MTF.A</p>	<p>หมายถึง ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ ตรวจให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์</p>
<p>ข้อสอบ MTF.P</p>	<p>หมายถึง ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกที่ ตรวจให้คะแนนแก่ความรู้บางส่วน</p>
<p>แบบสอบ T.MC</p>	<p>หมายถึง แบบสอบที่ประกอบด้วยข้อสอบ T.MC</p>
<p>แบบสอบ MTF.A</p>	<p>หมายถึง แบบสอบที่ประกอบด้วยข้อสอบ MTF.A</p>
<p>แบบสอบ MTF.P</p>	<p>หมายถึง แบบสอบที่ประกอบด้วยข้อสอบ MTF.P</p>
<p>ประสิทธิภาพของการตัดสินผลการเรียนเพื่อรู้จริงของแบบสอบ</p>	<p>หมายถึง อัตราการลดลงของการ เสี่ยงต่อการตัดสินผิดพลาด เมื่อใช้คะแนนจาก แบบสอบเป็นข้อมูลประกอบการตัดสินว่าผู้เรียน เป็นผู้รู้จริงหรือไม่รู้จริง ต่อการเสี่ยงต่อการ ตัดสินผิดพลาดเมื่อไม่ใช้คะแนนจากแบบสอบเป็น ข้อมูลประกอบการตัดสินเลย (ตัดสินโดยการสุ่ม)</p>

$$\eta = (R^* - R_0) / R^*$$

เมื่อ η คือ ดัชนีประสิทธิภาพของการตัดสินผลการเรียน
 R^* คือ ค่าฟังก์ชันความเสี่ยงเมื่อตัดสินโดยการสุ่ม

R_0 คือ ค่าฟังก์ชันความเสี่ยงเมื่อตัดสินโดยใช้ผล
การสอบ

ระดับความสามารถ

หมายถึง ระดับความรู้ความเข้าใจในเนื้อหาสาระวิชา
ศึกษา 251 วัดในหน่วยของ θ ที่มีค่าเฉลี่ยเท่า
กับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1
ในที่นี้จะแบ่งระดับความสามารถออกเป็น 5 ระดับ
คือ ระดับต่ำมาก $\theta \leq -2.5$
ระดับต่ำ $-2.5 < \theta \leq -1.0$
ระดับปานกลาง $-1.0 < \theta < +1.0$
ระดับสูง $+1.0 \leq \theta < +2.5$
ระดับสูงมาก $\theta \geq +2.5$

ผู้ทรงคุณวุฒิ

หมายถึง ผู้ที่มีคุณสมบัติดังต่อไปนี้

1. เป็นอาจารย์ผู้สอนในวิชาที่มีเนื้อหาเกี่ยวกับการประเมินผลการศึกษาและการสร้างแบบสอบ ในสถาบันอุดมศึกษา
2. สำเร็จการศึกษาไม่ต่ำกว่าชั้นปริญญาโท สาขาวิชาการศึกษาและประเมินผล สถิติ หรือวิจัยการศึกษา
3. เคยทำการสอนในวิชาดังกล่าวในข้อ 1 มาไม่น้อยกว่า 3 ปี

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ข้อความรู้ที่ได้จากการวิจัยครั้งนี้จะเป็นประโยชน์ต่อการตัดสินใจเลือกใช้ข้อสอบชนิดที่มีประสิทธิภาพสูงในการวัดผลสัมฤทธิ์ทางการเรียน ทั้งในแง่ของการใช้ผลสอบประมาณค่าความสามารถที่แท้จริง และการใช้ผลสอบตัดสินผลการเรียนเพื่อรู้จริง
2. ผลจากการวิจัยนี้ จะทำให้ได้ข้อความรู้เกี่ยวกับประสิทธิภาพข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก ซึ่งมองจากแนวคิดของทฤษฎีการตอบสนองต่อข้อสอบ ข้อความรู้ที่ได้นี้จะช่วยเสริมข้อความรู้เกี่ยวกับเรื่องนี้ที่มีอยู่ให้สมบูรณ์และชัดเจนยิ่งขึ้น