

การพัฒนาวิธีการตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์ Development of Computerized Scoring Algorithm for Constructed Response Item

สิริรัตน์ วิภาสศิลป์*

ผู้ที่ติดตามข่าวการร่างพระราชบัญญัติการศึกษาแห่งชาติฉบับใหม่จะเห็นความพยายามของผู้ที่มีส่วนเกี่ยวข้องทุกฝ่ายในการผลักดันให้มีการปฏิรูปการศึกษา ซึ่งเชื่อว่าเป็นวิถีทางนำไปสู่การผลิตพลเมืองที่มีคุณภาพของประเทศ ถ้าความพยายามดังกล่าวดำเนินไปอย่างมีประสิทธิภาพ ทุกชั้นตอนคงได้เห็นผลในไม่ช้า แนวคิดในการปฏิรูปการศึกษาในหลาย ๆ ประเทศเกิดจากความไม่พอใจในสภาพการจัดการศึกษาที่กำลังดำเนินอยู่ ผลผลิตของการศึกษาที่คิดไม่เป็น แก้ปัญหาต่าง ๆ ไม่ได้ ไม่มีความรู้ความสามารถในระดับที่พึงประสงค์ การเรียนการสอนทุกระดับในปัจจุบันส่วนใหญ่ยังเน้นการบรรยายของครู ผู้เรียนฟัง จด ท่อง แล้วจำเนื้อหาไปทำแบบทดสอบแบบเลือกตอบ และสิ่งของที่เรียนไปเมื่อเวลาผ่านไประยะหนึ่ง ซึ่งไม่เป็นผลดีแก่ทุกฝ่าย ในยุคโลกาภิวัตน์ ประเทศต่าง ๆ ต้องแข่งขันกันทุกด้าน สังคมจึงต้องการคนรุ่นใหม่ที่มีลักษณะคล่องแคล่ว มีวินัยในตนเอง มีความคิดเป็นระบบ ปรับตัวเก่ง มีความสามารถในการสื่อสาร มีจริยธรรม มีความรู้ลึกซึ้งในศาสตร์วิชาชีพ และมีความรู้กว้างขวางหลายศาสตร์ รู้จักวิเคราะห์ มีวิจารณ์ญาณ สนใจใฝ่รู้ใฝ่เรียน และมีทักษะการใช้เทคโนโลยี เป็นต้น เมื่อความต้องการของสังคมเปลี่ยนแปลงไป หน่วยงานที่หน้าที่รับผิดชอบจึงต้องจัดการศึกษาให้สอดคล้องกับความต้องการของสังคม การเรียนการสอนทุกระดับชั้นต้องเปลี่ยนแปลงไปเป็นการเรียนการสอนที่เน้นผู้เรียน ผู้เรียนเปลี่ยนบทบาทจากการฟังมาเป็น การแสวงหาความรู้ด้วยตนเอง ร่วมทำกิจกรรม และสร้างวินัยในการเรียนด้วยตนเอง เมื่อกระบวนการเรียนการสอนเปลี่ยนแปลงไป กระบวนการวัดและประเมินผลย่อมต้องเปลี่ยนแปลงตามไปด้วย แบบทดสอบชนิดเลือกตอบ ซึ่งเป็นเครื่องมือวัดผลสัมฤทธิ์ทางการเรียนที่นิยมใช้กันมานานนั้น มีข้อสงสัยว่าคะแนนจากการทำแบบทดสอบสามารถแสดงผลพัฒนาการของผู้เรียนได้หรือไม่ นอกจากนั้น เนื้อหาของแบบทดสอบโดยทั่วไปวัดทักษะพื้นฐานที่สังเกตได้โดยตรงจนมองข้ามความคิดขั้นสูง ความคิดที่ซับซ้อน และวิธีการแก้ปัญหาที่สอดคล้องกับการเรียนรู้ในโลกของความเป็นจริง จึงเกิดแนวคิดของการวัดประเมินผลแนวใหม่ (Alternative Assessment) ที่มุ่งวัดและประเมินความสามารถของผู้เรียนจากการปฏิบัติงาน มีการใช้ข้อสอบที่มุ่งให้ผู้เรียนค้นหาคำตอบด้วยตนเอง (constructed response item) หรือข้อสอบแบบเขียนตอบมากขึ้น

ข้อสอบที่มุ่งให้ผู้เรียนค้นหาคำตอบด้วยตนเอง หรือข้อสอบแบบเขียนตอบที่กล่าวถึงในบทความนี้มีความหมายกว้างมาก เริ่มตั้งแต่ปัญหาที่มุ่งให้ผู้เรียนเขียนตอบสั้น ๆ เขียนตอบแบบความเรียง ไปจนถึงการเสนอโครงการ และการสร้างผลงาน โดยใช้ทักษะหลายอย่างร่วมกัน ซึ่งเชื่อว่าจะทำให้

* ผู้ช่วยศาสตราจารย์ระดับ 8 ศูนย์วิจัยและพัฒนาแบบทดสอบ สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

ผลที่ได้จากการวัดมีความหมายมากขึ้น เนื่องจากสามารถวัดทักษะในการคิดระดับสูง เช่น การแก้ปัญหา (Problem solving) การคิดอย่างมีเหตุผล (reasoning) หรือการคิดอย่างพินิจเคราะห์ (critical thinking) ซึ่งสอดคล้องกับแนวโน้มของการวัดและประเมินผลการศึกษาในศตวรรษที่ 21 ซึ่งสอดคล้องกับแนวโน้มของการวัดและประเมินผลการศึกษา ตามที่ Hambleton และ Sireci¹ (1997) กล่าวไว้ว่า สิ่งที่มีอิทธิพลต่อการเปลี่ยนแปลงการจัดการศึกษารวมถึงการวัดและประเมินผลการศึกษาได้แก่ บริบทของสังคม แนวคิดของนักการศึกษาและผู้มีอำนาจในการกำหนดนโยบายด้านการศึกษา ความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ ทฤษฎีและแนวคิดใหม่ ๆ ทางการศึกษา การพัฒนากระบวนการคิดและวิธีวัดด้านจิตวิทยา โดยเฉพาะการวัดและประเมินผลการศึกษาในอนาคตมีลักษณะดังนี้

1. มีการใช้รูปแบบข้อสอบที่ผู้สอบต้องค้นหาคำตอบเอง (constructed response item format) มากขึ้น

2. เน้นการวัดทักษะกระบวนการคิดระดับสูง (higher - order cognitive skills)

3. ให้ความสำคัญกับการวัดและการแปลความหมายของคะแนนแบบอิงเกณฑ์มากขึ้น

4. มีรูปแบบการวัดที่มีความก้าวหน้าทันสมัย (sophistication of measurement models)

ซึ่งมีลักษณะดังนี้

4.1 ประเมินค่าความสามารถของผู้สอบจากผลการปฏิบัติงานอย่างมีอิสระโดยไม่มียึดติดกับคะแนนจากแบบทดสอบฉบับใดฉบับหนึ่ง

4.2 ให้คะแนนข้อสอบรายข้อแบบหลายค่า ซึ่งแตกต่างจากการให้คะแนนรายข้อแบบสองค่าคือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน

4.3 แปลความหมายความสามารถของผู้สอบ โดยเปรียบเทียบคะแนนจากสเกลร่วม (common scale)

4.4 ระบุแหล่งความคลาดเคลื่อนของคะแนนจากการวัดไว้หลายแหล่ง เช่น ผู้ตรวจให้คะแนน (rater) จำนวนชิ้นงาน (tasks) และวิธีการทดสอบ (mode of testing)

5. มีการเข้มงวดเกี่ยวกับความเที่ยงตรง (validity) และความเชื่อมั่น (reliability) ของการอ้างอิงคะแนนที่ได้จากการวัดมากขึ้น

6. มีการนำความก้าวหน้าของวิทยาการคอมพิวเตอร์มาใช้ในการพัฒนาแบบทดสอบ บริหารการสอบ ตรวจให้คะแนน และการรายงานผลคะแนนมากยิ่งขึ้น

สำหรับประเด็นของการนำความก้าวหน้าของวิทยาการคอมพิวเตอร์มาใช้ในการวัดและประเมินผลศึกษานั้น สามารถทำได้หลายประการ เช่น การจัดทำต้นฉบับแบบทดสอบ การจัดทำคลังข้อสอบ ซึ่งเครื่องคอมพิวเตอร์ช่วยอำนวยความสะดวกในการจัดทำต้นฉบับแบบทดสอบโดยการ

¹ Ronald K. Hambleton and Stephen G. Sireci, "Future Directions for Norm- Referenced and Criterion - Referenced Achievement Testing", **International Journal of Education Research**, 27(5), 1997, P.38.

คัดเลือกข้อสอบที่บรรจุไว้ในคลังตามเกณฑ์ที่ผู้ออกข้อสอบกำหนด โดยมีจำนวนข้อและพฤติกรรมที่ ต้องการวัดอย่างเหมาะสม รวมทั้งคำนวณค่าใช้จ่ายในการจัดทำแบบทดสอบแต่ละฉบับด้วย โปรแกรมสำเร็จรูปที่ใช้เพื่อการนี้ได้แก่ Con Test ซึ่งพัฒนาโดย Van der Linden และคณะ(1996) หรือ โปรแกรมอื่นๆ ช่วยให้สามารถบริหารการสอบได้อย่างสะดวกเป็นอย่างยิ่ง เมื่อมีความพร้อมผู้ สอบอาจแจ้งความประสงค์ขอสอบในเวลาที่ต้องการ แล้วทำข้อสอบบนเครื่องคอมพิวเตอร์ด้วยระบบ ออนไลน์(on line) และรู้ผลในทันทีทันใด นอกจากนั้นเครื่องคอมพิวเตอร์ยังมีประสิทธิภาพสูง สามารถใช้ในการวัดทักษะได้อย่างหลากหลายและปรับให้เหมาะสมกับความสามารถของแต่ละบุคคล ซึ่งกำลังเป็นที่นิยมกันทั่วไปในประเทศที่พัฒนาแล้ว คือ Computerized Adaptive Testing (CAT) ปัจจุบันคอมพิวเตอร์มีความสามารถในการตรวจให้คะแนนแบบทดสอบทั้งแบบเลือกตอบ และแบบ ที่ต้องคิดหาคำตอบด้วยตนเองในลักษณะต่าง ๆ แม้แต่การออกแบบอาคารของสถาปนิก (Bejar & Brann, 1994; อ้างจาก Hambleton, 1997) จากตัวอย่างที่กล่าวมาข้างต้นจะเห็นว่าในอนาคต คอมพิวเตอร์จะเข้ามามีบทบาทในการวัดและประเมินผลการศึกษามากยิ่งขึ้น นักวัดผลและผู้เกี่ยวข้อง จึงพยายามศึกษาค้นคว้าและทำการวิจัยเพื่อหาวิธีนำเทคโนโลยีคอมพิวเตอร์ไปพัฒนากระบวนการวัดและประเมินผลการศึกษาที่เหมาะสมกับศตวรรษใหม่ ดังตัวอย่างงานวิจัยเกี่ยวกับการพัฒนา วิธีการให้คะแนนโดยใช้เครื่องคอมพิวเตอร์เพื่อแทนที่การให้คะแนนโดยผู้เชี่ยวชาญในการวัดและ ประเมินการปฏิบัติโดย Brain E. Clauser และคณะ¹ โดยกล่าวถึงความเป็นมาและความสำคัญของ ปัญหา ดังนี้ การวัดและประเมินการปฏิบัติที่ผู้เชี่ยวชาญตรวจให้คะแนนผลงานของผู้สอบเป็นราย บุคคลนั้น ในกรณีที่มิได้มีผู้เข้าสอบจำนวนมากจะไม่สามารถดำเนินการได้อย่างมีประสิทธิภาพ เช่น ใน สหรัฐอเมริกาการสอบเพื่อออกหนังสือรับรองการประกอบอาชีพทางการแพทย์ ซึ่งมีผู้เข้าสอบมากกว่า 20,000 คน นอกจากนั้นการให้คะแนนโดยผู้เชี่ยวชาญต้องเสียค่าใช้จ่ายสูงและมีความคลาดเคลื่อนเกิดขึ้นได้มากกว่าการให้คะแนนโดยเครื่องคอมพิวเตอร์ ซึ่งกำหนดรูปแบบการตรวจให้ คะแนนไว้อย่างแน่นอนแล้ว ปัญหาดังกล่าวข้างต้นทำให้นักวิจัยพยายามพัฒนาระบบการวัดและ ประเมินการปฏิบัติ โดยใช้เครื่องคอมพิวเตอร์ และประสบผลสำเร็จเป็นที่ยอมรับได้ในการตรวจให้ คะแนนวิชาต่าง ๆ เช่น คณิตศาสตร์ (Bennett & Sebrechts, 1994 ; Sebrechts, Bennett, & Rock, 1991) สถาปัตยกรรม (Bejar, 1991) การเขียนโปรแกรมคอมพิวเตอร์ (Bruan, Bennett, Frye & Soloway, 1990) และวิธีให้การรักษาผู้ป่วย (Clauser, et al., 1995) ผลการวิจัยข้างต้นพบว่า การตรวจให้คะแนนโดยคณะผู้เชี่ยวชาญ และการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์มี สหสัมพันธ์สูงอย่างสมเหตุสมผล คณะผู้วิจัยจึงสนใจศึกษาเพื่อพัฒนาโปรแกรมการตรวจให้คะแนน การทดสอบทักษะการทํากายภาพบำบัดด้วยเครื่องคอมพิวเตอร์ ซึ่งวิธีการตรวจให้คะแนนที่ต้องการ พัฒนาขึ้น พิจารณาให้คะแนนผลการปฏิบัติตามความสำคัญของแต่ละขั้นตอนตามลำดับ พร้อมทั้ง

¹ Brain E. Clauser and Others, "Development of a Scoring Algorithm to Replace Rating for Scoring a Complex Performance - Based Assessment", *Applied Measurement in Education*, 10(4), 1997, P.345-358.

พิจารณาเวลาที่ผู้สอบใช้ในแต่ละขั้นตอนประกอบด้วย

การศึกษาครั้งนี้ The National Board of Medical Examiners ของสหรัฐอเมริกา มุ่งพัฒนาวิธีการตรวจให้คะแนนทักษะการทำกายภาพบำบัดของผู้สอบด้วยเครื่องคอมพิวเตอร์ เพื่อนำผลสอบไปใช้ในการพิจารณาออกหนังสือรับรองการประกอบอาชีพทางการแพทย์ในอนาคต โดยให้ผู้สอบทำแบบทดสอบในสถานการณ์จำลองเหมือนจริงบนเครื่องคอมพิวเตอร์ (computer-based simulation) ให้ผู้สอบแสดงวิธีการบำบัดเพื่อรักษาผู้ป่วย การทดสอบครั้งนี้ใช้เนื้อหาเป็นหลัก (text based) แต่มีสื่ออย่างอื่นประกอบด้วย ได้แก่ ภาพกราฟิก เสียง และวิดีโอ หรือใช้สื่อผสมทั้ง 3 อย่าง เพื่อสร้างสถานการณ์ให้สมจริง การทดสอบแต่ละรายการเริ่มจากการให้ผู้สอบชมภาพซึ่งนำเสนอจากเหตุการณ์เรื่องราวต่าง ๆ โดยมีการบรรยายประกอบ เช่น “ผู้ป่วยชายอายุ 45 ปี มีปัญหาเรื่องการปวดหลังมาพบท่าน นั่งรถที่ห้องตรวจโรคมีอาการปวดมากจนหาทำนั้งที่สบายไม่ได้.....” หลังจากฟังการบรรยายจบแล้ว ผู้สอบพิมพ์ข้อความเพื่อขอคู่มือวิธีการตรวจ การบำบัด และรายละเอียดอื่น ๆ ภายในเวลาที่กำหนด เงื่อนไขของผู้ป่วยจะเปลี่ยนแปลงไปตามปัญหาของผู้ป่วยและการปฏิบัติเพื่อแก้ไขปัญหาของผู้สอบ กำหนดระบบการทำงานให้เครื่องคอมพิวเตอร์เข้าใจค่าประมาณ 9,000 คำ ประกอบด้วยคำย่อ ชื่อทั่วไป ศัพท์เทคนิคที่ใช้ในการตรวจสอบ วินิจฉัย การให้ยา การให้คำปรึกษา และการบำบัด ผู้สอบมีโอกาสทำข้อสอบต่อไป (หรือหมดโอกาส) ขึ้นอยู่กับประสิทธิภาพในการวินิจฉัยและการแก้ไขปัญหาต่าง ๆ ภายในเวลาที่กำหนด ระหว่างที่ผู้สอบกำลังปฏิบัติตามขั้นตอนต่าง ๆ จะมีการบันทึกผลการปฏิบัติของผู้สอบในขั้นตอนต่าง ๆ ไว้ รวมทั้งเวลาที่ใช้ในแต่ละขั้นตอนด้วย เมื่อผู้สอบดำเนินการจนเสร็จเรียบร้อยแล้ว จะได้รับทราบบันทึกผลการปฏิบัติของตนเอง (transaction list)

กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักศึกษาแพทย์ปีที่ 3 จำนวน 280 คน ที่ต้องผ่านการทดสอบปัญหาที่เป็นกรณีเฉพาะ 7 กรณี สำหรับการเป็นแพทย์ฝึกหัด ปัญหาทั้งหมดเป็นปัญหาที่แพทย์ฝึกหัดต้องเรียนรู้ และเป็นการสอบภาคบังคับเพื่อนำผลไปใช้ในการพิจารณาให้เกรด เมื่อทำการวิเคราะห์ข้อมูลจากบันทึกผลการปฏิบัติของผู้สอบ จำนวน 1,960 ชิ้น (280 คน x 7 กรณี) โดยให้คณะผู้เชี่ยวชาญจำนวน 5 คน เป็นผู้ตรวจให้คะแนนผลงานของผู้สอบ ผู้เชี่ยวชาญเป็นอาจารย์ในคณะที่ผู้สอบศึกษาอยู่จึงรู้จักนักศึกษากลุ่มตัวอย่าง การให้คะแนนผลงานของผู้สอบแต่ละชั้น คณะผู้เชี่ยวชาญใช้วิธีการลงมติเอกฉันท์ ก่อนให้คะแนนคณะผู้เชี่ยวชาญทบทวนรายการทดสอบแต่ละกรณี (case) และร่วมกันอภิปรายความหมายของช่วงคะแนนจนทุกคนเกิดความเข้าใจตรงกัน เช่น ให้ 9 คะแนนถ้านักศึกษาวินิจฉัยปัญหาของผู้ป่วยได้อย่างถูกต้องทั้งขั้นต้นและขั้นสูง แล้วให้การบำบัดที่เหมาะสมตามลำดับขั้น ส่วนการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ ให้คณะผู้เชี่ยวชาญชุดเดิมทำตามกระบวนการที่กล่าวมาแล้ว และกำหนดเกณฑ์ในการให้คะแนนแต่ละช่วงอย่างเป็นลายลักษณ์อักษร โดยพิจารณาเงื่อนไขต่าง ๆ ที่เกิดขึ้นอย่างรอบคอบ แล้วนำเกณฑ์การให้คะแนนที่คณะผู้เชี่ยวชาญสร้างขึ้นกำหนดกรอบความคิดเชิง ตรรกศาสตร์ ในการให้คะแนนผลงานของนักศึกษาด้วยเครื่องคอมพิวเตอร์ ความคิดพื้นฐานในการให้คะแนนขึ้นอยู่กับปฏิบัติที่เหมาะสมของผู้สอบ

ในแต่ละขั้นตอน คะแนนรวมจากการปฏิบัติในแต่ละขั้นตอนถูกปรับลดถ้ามีการปฏิบัติที่ไม่เหมาะสม และอาจทำให้เกิดอันตรายแก่ผู้ป่วย การปรับลดคะแนนอาศัยผลจากการอภิปรายของคณะผู้เชี่ยวชาญ เช่น ปรับลด 1 คะแนน ถ้าปฏิบัติไม่เหมาะสม 2-4 ครั้ง ปรับลด 2 คะแนน ถ้าปฏิบัติไม่เหมาะสมมากกว่า 4 ครั้ง ปรับลด 4 คะแนน ถ้าทำให้ผู้ป่วยได้รับอันตรายถึงกับเสียชีวิต การปรับลดคะแนนทำให้ผู้สอบอาจได้คะแนนติดลบได้ด้วย การแปลงความเห็นของคณะผู้เชี่ยวชาญมาเป็นเกณฑ์การตรวจให้คะแนนผลการปฏิบัติของผู้สอบด้วยเครื่องคอมพิวเตอร์เป็นสิ่งที่ซับซ้อนยุ่งยากและอาจทำให้เกิดความคลาดเคลื่อนขึ้นได้ ในการศึกษาครั้งนี้จึงตรวจสอบความถูกต้องของการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ โดยสุ่มนักศึกษาในกลุ่มตัวอย่างแบ่งเป็น 2 กลุ่มย่อยเท่า ๆ กัน นำคะแนนของผู้สอบในกลุ่มย่อยที่หนึ่งมาคำนวณความแตกต่างระหว่างการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์และการตรวจให้คะแนนโดยคณะผู้เชี่ยวชาญ ชิ้นงานที่มีความแตกต่างของคะแนนมากที่สุดจะได้รับการวิเคราะห์เพื่อหาข้อบกพร่องของระบบการตรวจให้คะแนนด้วยคอมพิวเตอร์แล้วทำการตรวจซ้ำเพื่อปรับคะแนนให้เหมาะสมยิ่งขึ้น ส่วนคะแนนของกลุ่มย่อยที่สองที่ได้จากการตรวจด้วยเครื่องคอมพิวเตอร์ไม่มีการปรับคะแนน

การวิจัยครั้งนี้ตรวจสอบความเหมาะสมของให้คะแนนด้วยเครื่องคอมพิวเตอร์ดังนี้

1) ตรวจสอบระดับความสอดคล้องของคะแนนที่ได้จากการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์และคะแนนที่ได้จากการตรวจให้คะแนนโดยคณะผู้เชี่ยวชาญ 2) ตรวจสอบการจำแนกผลงานของผู้สอบกลุ่มที่ปฏิบัติจนเกิดผลดีในการทำกายภาพบำบัด กับกลุ่มที่ทำให้ผู้ป่วยเกิดความเสียหายหรือมีอันตราย (จำแนกกลุ่มผ่าน-ไม่ผ่าน) การตัดสินผลงานผ่าน-ไม่ผ่านให้เทียบกับเกณฑ์ของผู้เชี่ยวชาญอิสระอีกกลุ่มหนึ่งซึ่งพิจารณาผลงานของผู้สอบรายการละ 40 ชิ้น และตัดสินโดยใช้การลงมติเอกฉันท์

ตารางที่ 1 แสดงสัมประสิทธิ์สหสัมพันธ์ระหว่างการให้คะแนนโดยคณะผู้เชี่ยวชาญและการให้คะแนนโดยคำนวณจากเครื่องคอมพิวเตอร์

รายการ	ก่อนปรับ ทั้งหมด	ภายหลังการปรับ		
		กลุ่มย่อย 1	กลุ่มย่อย 2	ทั้งหมด
1	.80	.85	.81	.83
2	.71	.78	.86	.82
3	.66	.81	.73	.78
4	.66	.65	.74	.70
5	.40	.65	.67	.66
6	.79	.81	.83	.80
7	.32	.78	.70	.74

* กลุ่มย่อย 1 เป็นกลุ่มที่มีการปรับคะแนน * กลุ่มย่อย 2 เป็นกลุ่มที่ไม่มีการปรับคะแนน

ผลการวิจัยพบว่า การตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์และการตรวจให้คะแนนโดยคณะผู้เชี่ยวชาญมีความสอดคล้องกันค่อนข้างสูง ดังแสดงในตารางที่ 1

มีการเปรียบเทียบผลจากการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์และการตรวจให้คะแนนด้วยผู้เชี่ยวชาญในหลายลักษณะพบว่า การตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์สามารถดำเนินการได้อย่างมีประสิทธิภาพ ถ้ามีการวางแผนและได้รับความร่วมมือจากผู้ที่เกี่ยวข้องทุกฝ่ายเป็นอย่างดีโดยเฉพาะผู้เชี่ยวชาญด้านคอมพิวเตอร์ สำหรับการวิจัยครั้งนี้มีข้อจำกัดเกี่ยวกับจำนวนกลุ่มตัวอย่าง รายการที่ทดสอบ และผู้เชี่ยวชาญซึ่งมีจำนวนไม่มากนัก นอกจากนี้การที่ผู้เชี่ยวชาญรู้จักกับผู้สอบทำให้การตรวจให้คะแนนเกิดความลำเอียงได้ จึงควรทำการวิจัยซ้ำโดยออกแบบการวิจัยให้รัดกุมยิ่งขึ้นเพื่อจะได้นำผลที่ได้จากการวิจัยไปประยุกต์ใช้ได้อย่างกว้างขวางต่อไป เนื่องจากการตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์มีความคงเส้นคงวามีความยุติธรรมกับผู้สอบทุกคน และการตรวจให้คะแนนสามารถดำเนินการได้อย่างรวดเร็วทันต่อความต้องการใช้ผลสอบ

จากตัวอย่างงานวิจัยที่กล่าวมานั้นจะเห็นว่าเทคโนโลยีคอมพิวเตอร์ช่วยอำนวยความสะดวกในการตรวจให้คะแนนการตอบข้อสอบแบบเขียนตอบได้อย่างมีประสิทธิภาพ ผู้ที่สนใจอาจนำมาเป็นแนวทางในการศึกษาค้นคว้าเพื่อพัฒนาวิธีการตรวจให้คะแนนการตอบข้อสอบแบบเขียนตอบในระดับการศึกษาต่าง ๆ ได้เป็นอย่างดี เช่น ในการทดสอบซึ่งมีผู้เข้าสอบเป็นจำนวนมาก สำหรับวิชาที่ใช้ข้อสอบแบบเขียนตอบนั้นต้องใช้ผู้เชี่ยวชาญตรวจให้คะแนนเป็นจำนวนมากและใช้เวลาในการตรวจนานและยังมีปัญหาเกี่ยวกับความเป็นปรนัยในการตรวจให้คะแนนด้วย หากสามารถพัฒนาวิธีการตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์จะทำให้การจัดการศึกษาในระบบการเรียนการสอนทางไกลมีความก้าวหน้าทันยุคทันสมัยยิ่งขึ้น แม้แต่ในระดับมัธยมศึกษาก็อาจพัฒนาวิธีการตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์ได้เช่นกัน จากการติดตามการศึกษาเพื่อทำวิทยานิพนธ์ของนักศึกษาระดับมหาบัณฑิตพบว่ามีการศึกษาเพื่อพัฒนาบทเรียนวิชาต่าง ๆ โดยใช้คอมพิวเตอร์ช่วยสอนในหลายวิชา ในการศึกษาเรื่องดังกล่าว ผู้วิจัยใช้แบบทดสอบปรนัยชนิดเลือกตอบเป็นเครื่องมือวัดผลสัมฤทธิ์ทางการเรียนเมื่อสิ้นสุดการทดลอง หากมีการศึกษาต่อไปโดยใช้แบบทดสอบแบบเขียนตอบ และตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ เนื่องจากผู้เรียนมีความคุ้นเคยกับการศึกษาบทเรียนโดยใช้คอมพิวเตอร์อยู่แล้ว จะทำให้การเรียนการสอนมีประสิทธิภาพยิ่งขึ้น เพราะสามารถตรวจสอบทักษะการคิดระดับสูงของผู้เรียนได้มากขึ้น วิธีการในการศึกษาอาจใช้กลุ่มตัวอย่างในโรงเรียนที่มีผู้วิจัยเกี่ยวกับการใช้คอมพิวเตอร์ช่วยสอนวิชาใดวิชาหนึ่งอยู่แล้ว สมมุติว่าเป็นวิชาคณิตศาสตร์เรื่องสถิติสำหรับนักเรียนชั้นมัธยมศึกษา ปีที่ 3 อาจกำหนดแนวทางในการพัฒนาแบบทดสอบและวิธีการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ดังนี้

แต่งตั้งคณะกรรมการสร้างและพัฒนาแบบทดสอบสำหรับตอบและตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ คณะกรรมการควรประกอบด้วยผู้เชี่ยวชาญด้านเนื้อหา ผู้เชี่ยวชาญด้านคอมพิวเตอร์และนักวัดผล หน้าที่ของคณะกรรมการได้แก่ ตรวจสอบตารางวิเคราะห์หลักสูตรของวิชาดังกล่าว นำมากำหนดสถานการณ์เพื่อสร้างโจทย์ให้ผู้สอบหาคำตอบด้วยตนเองให้ครอบคลุมตาม

วัตถุประสงค์และเนื้อหาที่ระบุไว้ในตารางวิเคราะห์หลักสูตร พร้อมทั้งกำหนดแนวตอบที่มีความเป็นไปได้ทั้งหมดพร้อมทั้งกำหนดเกณฑ์ในการให้คะแนนคำตอบต่าง ๆ ไว้อย่างชัดเจน เมื่อดำเนินการแล้วให้ผู้เชี่ยวชาญด้านเนื้อหาที่กำหนดขึ้นอีกชุดหนึ่งช่วยตรวจสอบความเที่ยงตรงเชิงเนื้อหา ความเหมาะสมของแนวตอบและเกณฑ์การให้คะแนน หลังจากนั้นคณะกรรมการชุดเดิมปรับปรุงคำถามแนวตอบตามข้อเสนอแนะของผู้เชี่ยวชาญด้านเนื้อหา (ชุดใหม่) อย่างรอบคอบจนคณะกรรมการเห็นชอบพร้อมกันแล้วจึงพัฒนาโปรแกรมสำเร็จรูปในการสร้างคำถาม การตอบคำถาม และการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ แก้ไขปรับปรุงโปรแกรมให้มีความสะดวกต่อการนำไปใช้แล้วทดลองใช้กับกลุ่มตัวอย่าง นำผลที่ได้มาแก้ไขปรับปรุงจนกว่าสามารถดำเนินการได้อย่างมีประสิทธิภาพสูงสุด อาจใช้วิธีการตรวจสอบผลของการตรวจให้คะแนนด้วยเครื่องคอมพิวเตอร์ตามแนวทางที่กล่าวมาแล้ว ในรายงานการวิจัยข้างต้นพร้อมทั้งสอบถามความคิดเห็นของผู้ที่เกี่ยวข้องทุกฝ่ายเพื่อนำความเห็นเหล่านั้นมาประกอบการปรับปรุงแก้ไขระบบให้สามารถนำไปปฏิบัติได้จริง

การพัฒนาวิธีตรวจให้คะแนนข้อสอบแบบเขียนตอบด้วยเครื่องคอมพิวเตอร์ไม่ใช่เรื่องง่าย ต้องอาศัยผู้ที่มีความรู้ความสามารถ มีประสบการณ์มีความตั้งใจและมีเวลาศึกษาค้นคว้าอย่างจริงจัง นอกจากนั้นยังต้องมีการสนับสนุนทั้งกำลังใจและกำลังทรัพย์จากผู้ที่เกี่ยวข้องจึงจะทำให้การศึกษาค้นคว้าประสบความสำเร็จได้ ผู้เขียนหวังว่าบทความนี้คงจุดประกายความคิดและมีผู้สนใจนำไปศึกษาค้นคว้าอย่างเป็นรูปธรรมต่อไป จนสามารถประยุกต์ใช้เป็นเรื่องมีอรรถผลการศึกษาในสถานศึกษาทุกระดับ หรือใช้เป็นเครื่องมือวัดผลในโอกาสต่าง ๆ เช่น การสอบคัดเลือกบุคคลเข้าศึกษาต่อในระดับอุดมศึกษา การสอบคัดเลือกบุคคลเข้าทำงาน การทดสอบเพื่อขอรับใบประกอบวิชาชีพต่าง ๆ เป็นต้น ถือว่าเป็นการส่งเสริมให้ผู้สอบมีโอกาสแสดงทักษะในการคิดขั้นสูง เพื่อหาแนวทางแก้ปัญหาในสถานการณ์ต่าง ๆ ได้ ซึ่งสอดคล้องกับแนวคิดในการพัฒนาคนตามหลักของการปฏิรูปการศึกษา

บรรณานุกรม

- เกษม พิงพา, การพัฒนาบทเรียนคอมพิวเตอร์ช่วยสอนวิชาคณิตศาสตร์เรื่องความน่าจะเป็นเบื้องต้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนวัดราชบพิตร กรุงเทพมหานคร. วิทยานิพนธ์หลักสูตรปริญญา ศึกษาศาสตรมหาบัณฑิต แขนงวิชาหลักสูตรและการสอน สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2541.
- จินตนา ธนวิบูลย์ชัย, “การวัดการปฏิบัติ สภาพปัญหาและแนวทางการพัฒนา”, วารสาร วัดผลการศึกษา. พ.ศ. - ส.ศ., 2536.

- ชัยฤทธิ คีลาเดช, การพัฒนาแฟ้มสะสมงานในการประเมินผลการเรียนวิชาภาษาอังกฤษระดับ
ชั้นมัธยมศึกษาปีที่ 3. ปรินญาณิพนธ์ดุขฎิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มหาวิทยาลัยศรีนครินทรวิโรฒ, 2540.
- บุญเชิด ภิญโญอนันตพงษ์, “Portfolio Assessment”, วารสารวัดผลการศึกษา. 17 (49),
พ.ค. - ส.ค., 2538.
- ศิริวรรณ ตรีพงษ์พันธุ์, การพัฒนาบทเรียนวิชาคณิตศาสตร์เรื่องสถิติโดยใช้คอมพิวเตอร์
ช่วยสอน สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนสุราษฎร์พิทยา จังหวัด
สุราษฎร์ธานี. วิทยานิพนธ์ตามหลักสูตร ปรินญาณิศึกษาศาสตรมหาบัณฑิต แขนงวิชา
หลักสูตรและการสอน สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช, 2538.
- สิริรัตน์ วิภาสศิลป์, “ทางเลือกในการวัดและประเมินผลการศึกษาของประเทศไทย : การวัดและ
ประเมินการปฏิบัติ” วารสารสุโขทัยธรรมาธิราช. 11 (3), กันยายน - ธันวาคม, 2541.
- สุวิมล ว่องวานิช, การวัดทักษะการปฏิบัติ. จุลสารการทดสอบอันดับที่ 4, ศูนย์ทดสอบทาง
การศึกษาร่วมกับศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์
มหาวิทยาลัย, กทม., 2539.
- สมหวัง พิธิยานุวัฒน์, วิถีวิทยาการประเมินทางการศึกษา. ภาควิชาวิจัยทางการศึกษา
คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย, 2540.
- อุทุมพร (ทองอุไทย) จามรมาน, การตีค่าความสามารถที่แท้จริงของผู้เรียนเพื่อการปฏิรูป
การศึกษา. หจก. ฟันนี้พับบลิชชิ่ง, กทม., 2540.
- Anthony J.Nitko, **Educational Assessment of Students**. 2nd editon. USA. : Prentice Hall, Inc.,
1996.
- Brain E.Clauser and Others, “Development of a Scoring Algorithm to Replace Expert Rating
for Scoring a Complex Performance-Based Assessment”, **Applied Measurement in
Education**. 10 (4), 1997.
- Carl S.Parke and Suzanne Lane, “Learning from Performance Assessment in Math”,
Educational Leadership. 54 (4), 1997.
- Ronald K. Hambleton and Stephen G.Sireci, “Future Directions for Norm – Referenced and
Criterion – Referenced Achievement Testing”, **International of Education Research**.
27(5), 1997.