บทที่ 1
บทนำ

ความเป็นมาและความสำคัญ

ปัจจุบันเป็นยุคที่นานาประเทศมีการแข่งขันทางด้านเศรษฐกิจที่อาศัยความรู้สมัยใหม่มาสร้างความได้เปรียบในการแข่งขัน (competitive advantages) และความรู้สมัยใหม่ที่ช่วยในการแข่งขันส่วนใหญ่มาจากวิทยาศาสตร์และเทคโนโลยีที่พัฒนาขึ้น ประเทศใดที่เข้าถึงเทคโนโลยีได้มากก็ย่อมสร้างความได้เปรียบได้มาก ดังนั้นการส่งเสริมการศึกษาวิทยาศาสตร์ให้มีคุณภาพจึงควรมีความสำคัญเป็นลำดับแรกๆ เพื่อให้ประเทศไทยมีความเข้มแข็งในการแข่งขันทางเศรษฐกิจได้ การพัฒนาการศึกษาวิทยาศาสตร์จึงมีความจำเป็นอย่างยิ่ง กระทรวงศึกษาธิการในฐานะผู้รับผิดชอบการจัดการศึกษาของประเทศได้มีการปฏิรูปการจัดการศึกษา เริ่มขึ้นใน ปี พ.ศ. 2540 แต่ที่ผ่านมา 17 ปี ผลการศึกษาจำนวนมากเช่น การประเมินการรู้วิทยาศาสตร์ (scientific literacy) ของนักเรียนในโครงการการประเมินนักเรียนระดับนานาชาติ (PISA) จำนวนหลายครั้ง ชี้ว่าคุณภาพการเรียนการสอนวิทยาศาสตร์ระดับพื้นฐานของไทยยังไม่ดีโดยเฉพาะระดับมัธยมศึกษาตอนต้น ซึ่งเป็นนักเรียนที่ได้รับการจัดการเรียนการสอนระดับที่สามารถมีความรู้ สามารถออกไปประกอบอาชีพได้ การประเมินการรู้วิทยาศาสตร์ของ PISA ชี้ว่า นักเรียนไทยมีปัญหาในการนำความรู้วิทยาศาสตร์ไปใช้ หากไม่สามารถยกระดับการเรียนการสอนวิทยาศาสตร์ให้ดีขึ้นได้ การพัฒนาเศรษฐกิจและสังคมจะมีปัญหามาก เพราะวิทยาศาสตร์ไม่ใช่แค่ศาสตร์ที่ทำให้มีความรู้สำหรับการพัฒนาเทคโนโลยีสำหรับการพัฒนาประเทศเท่านั้น แต่ยังเป็นศาสตร์ที่ช่วยพัฒนาทักษะการคิด เหตุผล และการใช้ชีวิตในสังคมด้วย

การรู้วิทยาศาสตร์เป็นการเข้าใจหน้าที่ของวิทยาศาสตร์แบบกว้างๆ โดยเป็นเป้าหมายของการจัดการศึกษาทั่วๆ ไป ที่ไม่ใช่การเตรียมคนเพื่อการทำงานด้านวิทยาศาสตร์ (Lau, 2009) การรู้วิทยาศาสตร์จึงเป็นเป้าหมายสำคัญของการจัดการศึกษาวิทยาศาสตร์ เพื่อส่งเสริมการจัดการเรียนการสอนวิทยาศาสตร์ให้มีคุณภาพ โดยคาดหวังให้นักเรียนได้รับการพัฒนาสมรรถนะ และมีความรู้ด้านวิทยาศาสตร์อย่างลึกซึ้ง ประเทศไทยก็มีการพัฒนาและปรับปรุงการจัดการเรียนการสอนหลายๆ ด้าน โดยมีสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) เป็นหน่วยงานหลักในการพัฒนาครู หลักสูตรการสอนวิทยาศาสตร์ และมีสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์กรมหาชน) หรือ สทศ. เป็นหน่วยงานที่รับผิดชอบในการประเมินความรู้ของนักเรียนในทุกกลุ่มสาระการเรียนรู้ รวมทั้งวิชาวิทยาศาสตร์ด้วย รูปแบบการทดสอบที่ สทศ. ใช้เป็นการทดสอบวัดผลสัมฤทธิ์ทางการเรียน โดยจัดสอบผ่านแบบทดสอบธรรมดา ซึ่งนักการศึกษาและนักประเมินด้านวิทยาศาสตร์เชื่อว่าเป็นรูปแบบการทดสอบที่วัดความรู้ความจำของนักเรียนเป็นส่วนใหญ่ แต่ไม่สามารถวัดความรู้ขั้นสูงของนักเรียนได้ ทำให้ครูผู้สอนขาดข้อมูลเกี่ยวกับสมรรถนะและการรู้วิทยาศาสตร์ (scientific literacy) ของนักเรียนที่ถูกต้อง จึงเป็นปัญหาทำให้ครูพัฒนานักเรียนได้ไม่ตรงกับสมรรถนะและการรู้วิทยาศาสตร์ของนักเรียน

ปัจจุบันนี้ มีแนวโน้มที่ชัดเจนให้เห็นว่า การวัดและประเมินความรู้ สมรรถนะ และการรู้วิทยาศาสตร์ควรวัดทักษะขั้นสูงของนักเรียน เช่น ในสหรัฐอเมริกา มีการกล่าวถึงมาตรฐานด้าน

วิทยาศาสตร์ในอนาคต (Next Generation of Science Standards) มาตรฐานของความสำเร็จด้านวิทยาศาสตร์ (Standards for Science Success) และกรอบสำหรับการจัดการศึกษาวิทยาศาสตร์ระดับประถมศึกษาถึงมัธยมศึกษา (Framework for K-12 Science Education) เพื่อเป็นกรอบในการวัดสิ่งที่นักเรียนควรรู้และสามารถทำได้ ซึ่งเป็นกรอบของการเรียนการสอนวิทยาศาสตร์อย่างลึกซึ้ง และมีการเสนอให้นำเทคโนโลยีที่ทันสมัยมาใช้ในการพัฒนาการทดสอบวิทยาศาสตร์ทั้งในระดับนานาชาติ ระดับประเทศ และระดับเขตพื้นที่เพื่อให้สามารถวัดความรู้ และทักษะขั้นสูงของนักเรียนได้ (Quellmalz, Davenport, Timms, DeBoer, Jordan, Huang, Buckley, 2013) นักวิชาการและนักประเมินด้านการศึกษาวิทยาศาสตร์เห็นว่าเทคโนโลยีและคอมพิวเตอร์สามารถนำมาใช้ในการจัดสอบเพื่อแก้ไขข้อจำกัดของการทดสอบแบบธรรมดา และมีการนำมาใช้จริงแล้ว เช่น การทดสอบวัดความก้าวหน้าทางการศึกษาแห่งชาติของสหรัฐอเมริกา (National Assessment of Educational Progress หรือ NAEP) เริ่มมีการใช้เทคโนโลยีในการจัดสอบในปี ค.ศ. 2009 โครงการประเมินนักเรียนระดับนานาชาติ หรือ PISA ก็มีการทดสอบในรูปแบบของคอมพิวเตอร์ตั้งแต่ ปี ค.ศ. 2006 เป็นต้นมา การทดสอบความรู้และทักษะวิทยาศาสตร์ด้วยเทคโนโลยีเหล่านี้สามารถวัดความรู้ที่ลึกซึ้งได้เพราะสามารถจำลองเหตุการณ์เพื่อให้นักเรียนตอบดำเนินการตรวจสอบได้ ซึ่งเป็นลักษณะของการทดสอบจากการจำลองสถานการณ์ (science simulations) การทดสอบเหล่านี้ต้องมีการพัฒนาข้อสอบที่สามารถโต้ตอบกับผู้สอบได้ (interactive items) โดยใช้เทคโนโลยีที่ทันสมัยมาใช้ในการพัฒนาข้อสอบ (technology-enhanced item)

เทคโนโลยีทางการศึกษาที่ทันสมัยจะทำให้สามารถประเมินทักษะ ความรู้ สมรรถนะขั้นสูง รวมทั้งการรู้วิทยาศาสตร์ของนักเรียนได้ โดยการนำเทคโนโลยีมาใช้ในการพัฒนาข้อสอบที่แสดงเหตุการณ์จำลอง นักเรียนสามารถศึกษาเหตุการณ์ในสถานการณ์จำลองได้โดยคลิ๊กเมาส์ ทำให้เป็นการทดสอบที่เสมือนจริง และทำให้นักเรียนตื่นตัวในขณะทดสอบ การทดสอบแบบนี้ทำให้นักเรียนได้ทดลอง สังเกต ออกแบบการทดลอง และสามารถหาคำตอบได้ด้วย ผลการวิจัยจำนวนมากชี้ว่าการทดสอบผ่านสถานการณ์จำลองมีความเป็นไปได้ในการนำไปใช้ประเมินสมรรถนะด้านวิทยาศาสตร์ของนักเรียน กล่าวคือ นอกจากจะสามารถวัดความรู้ของนักเรียนได้แล้ว ยังช่วยพัฒนาความเข้าใจเนื้อหาของวิทยาศาสตร์ได้อย่างลึกซึ้ง และส่งเสริมทักษะการแก้ปัญหาทางวิทยาศาสตร์ (Adams et al., 2008; Buckey, Gobert, Howitz, & O'Dwyer, 2010) เพราะนักเรียนสามารถสร้างการตรวจสอบ ปรับปรุงการดำเนินงานตรวจสอบ จึงช่วยพัฒนาความรู้ได้มากกว่าการใช้แบบทดสอบธรรมดา

จากประโยชน์ของเทคโนโลยีด้านการทดสอบที่กล่าวมาข้างต้น คณะนักวิจัยจึงเห็นว่าการใช้เทคโนโลยีพัฒนาแบบทดสอบวิทยาศาสตร์เพื่อส่งเสริมการเรียนรู้วิทยาศาสตร์ในรูปแบบของการประเมินเพื่อการเรียนรู้ (assessment for learning) มีความจำเป็นต่อทั้งครูและนักเรียน การพัฒนาแบบทดสอบในรูปแบบที่นักเรียนสามารถเข้ามาทดสอบและศึกษาด้วยตนเองเวลาใดก็ได้จะเป็นการส่งเสริมการเรียนรู้วิทยาศาสตร์ได้มาก นอกจากนี้ครูยังสามารถนำไปใช้การจัดการเรียนการสอนได้ด้วย อันจะเป็นวิถีทางหนึ่งที่จะช่วยยกระดับคุณภาพการเรียนการสอนวิทยาศาสตร์ของประเทศให้ดีขึ้น เพื่อให้ผู้เรียนมีความรู้ และทักษะด้านวิทยาศาสตร์อย่างลึกซึ้ง อันจะส่งผลต่อการพัฒนาความสามารถในการแข่งขันของประเทศต่อไป

ดังนั้น ผู้วิจัยจึงมีความต้องการจะพัฒนาแบบทดสอบวิทยาศาสตร์ที่เป็นข้อสอบที่เป็นสถานการณ์จำลอง เพื่อให้นักเรียนสามารถโต้ตอบกับข้อมูลในสถานการณ์ที่กำหนดให้ได้ และพัฒนาระบบการจัดสอบวัดการรู้วิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต เพื่อนำไปให้นักเรียนระดับมัธยมศึกษาตอนต้น ได้ใช้ อันจะเกิดประโยชน์สองด้าน คือ 1) ส่งเสริมคุณภาพการศึกษาวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นของประเทศ เครื่องมือการทดสอบที่พัฒนาขึ้นจะมีประโยชน์นักเรียนทั้งประเทศ เพราะนักเรียนระดับมัธยมศึกษาตอนต้นทุกคนสามารถเข้ามาใช้ได้ผ่านทั้งทางอินเทอร์เน็ต แท็บเล็ต ตลอดจนโทรศัพท์มือถือ ทำให้เกิดการเรียนรู้และนักเรียนได้รับการพัฒนาทุกเมื่อ ทุกเวลา และ 2) เป็นการเตรียมความพร้อมของนักเรียนไทยในการประเมิน PISA ในอนาคต ซึ่งการประเมิน PISA ในอนาคตจะมีการทดสอบการรู้วิทยาศาสตร์ที่เป็นสมรรถนะขั้นสูงผ่านการทดสอบทางเทคโนโลยีสมัยใหม่

วัตถุประสงค์ของการวิจัย

การวิจัยนี้มุ่งพัฒนาระบบการประเมินการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ต เพื่อส่งเสริมการเรียนรู้ขั้นสูงในวิชาวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น โดยมีวัตถุประสงค์ย่อยดังนี้

1. เพื่อสร้างข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น
2. เพื่อพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต
3. เพื่อตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต
4. เพื่อนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต

ขอบเขตของการวิจัย

1. ระบบประเมินการรู้วิทยาศาสตร์ที่พัฒนาประกอบด้วยการพัฒนาข้อสอบการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น เทคโนโลยีการจัดสอบ และการทดสอบผ่านอินเทอร์เน็ต
2. เนื้อหาที่ต้องการนำมาพัฒนาแบบทดสอบครอบคลุมเนื้อหาตามหลักสูตรแกนกลางการศึกษา พุทธศักราช 2551
3. สมรรถนะทางวิทยาศาสตร์ที่จะนำมาใช้ในการประเมินใช้กรอบการประเมินของการประเมินนักเรียนระดับนานาชาติ (PISA) เพื่อวัดสมรรถนะขั้นสูงที่นานาชาติยอมรับ และเป็นการเตรียมนักเรียนให้พร้อมกับการทดสอบ PISA ในอนาคต
5. กลุ่มเป้าหมายของการพัฒนาครั้งนี้ คือ นักเรียนที่มีอายุ 15 ปี ซึ่งเป็นวัยที่สามารถนำความรู้ด้านวิทยาศาสตร์ไปใช้ในการศึกษาต่อและประกอบอาชีพ ตามกรอบการประเมินนักเรียนระดับนานาชาติ (PISA)

นิยามศัพท์

1. ระบบประเมินการรู้วิทยาศาสตร์ หมายถึง องค์ประกอบของการประเมินการรู้วิทยาศาสตร์ ประกอบด้วยระบบการนำเข้าข้อสอบสู่ระบบ ระบบการเขียน/ปรับข้อสอบ ระบบการสอบผ่านอินเทอร์เน็ต และ ระบบการรายงานผลการสอบ

2. การรู้วิทยาศาสตร์ หมายถึง ความรู้สามารถในการนำความรู้ทางวิทยาศาสตร์ไปใช้อธิบายปรากฏการณ์ที่เกี่ยวกับวิทยาศาสตร์ และการหาความรู้ด้วยกระบวนการวิทยาศาสตร์อย่างเหมาะสม

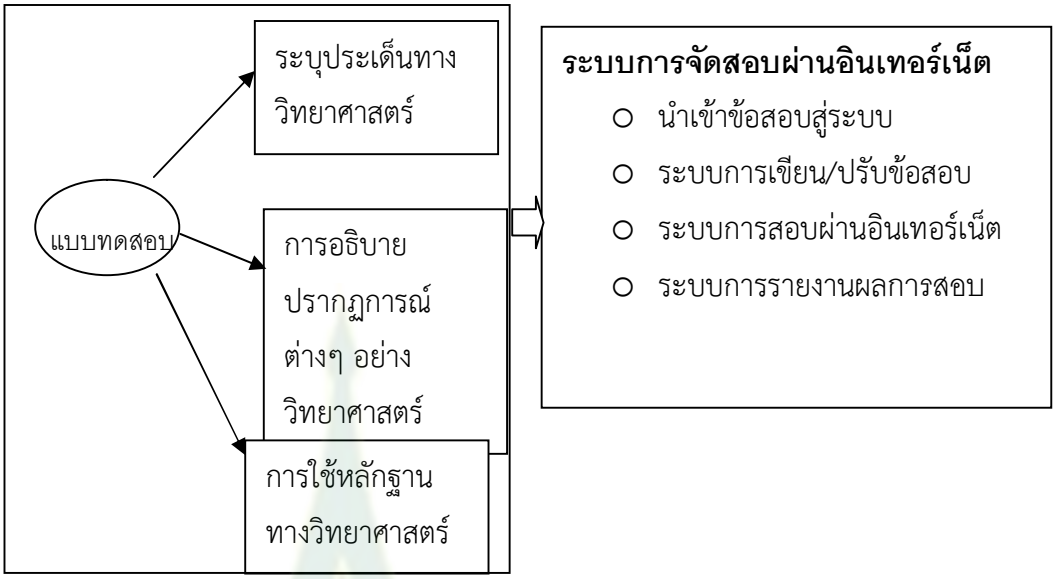
3. การระบุประเด็นทางวิทยาศาสตร์ หมายถึง ความสามารถในการระบุว่าประเด็นใดบ้างที่สามารถตรวจสอบได้อย่างวิทยาศาสตร์ การระบุคำสำคัญที่ต้องใช้ค้นหาข้อมูลทางวิทยาศาสตร์ และการรู้ลักษณะสำคัญของการตรวจสอบทางวิทยาศาสตร์

4. การอธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์ หมายถึง การใช้ความรู้วิทยาศาสตร์ในสถานการณ์ที่กำหนด การอธิบายหรือแปลผลปรากฏการณ์อย่างวิทยาศาสตร์ และทำนายการเปลี่ยนแปลง การบอกถึงการบรรยาย อธิบาย และทำนายที่เหมาะสม

5. การใช้หลักฐานทางวิทยาศาสตร์ หมายถึง การแปลผลหลักฐานทางวิทยาศาสตร์ สรุปและสื่อสารให้เข้าใจได้ การระบุข้อตกลงเบื้องต้น หลักฐาน และเหตุผลเบื้องหลังข้อสรุป และการสะท้อนนัยของการนำผลการพัฒนาทางวิทยาศาสตร์และเทคโนโลยีไปใช้ในสังคม

กรอบแนวความคิดของการวิจัย

กรอบแนวคิดในการวิจัยเป็นการบูรณาการการทดสอบและเทคโนโลยีเข้าด้วยกัน ซึ่งจะมีสองส่วนที่เกี่ยวข้องกัน คือ กรอบการสร้างข้อสอบวิทยาศาสตร์ และกรอบการพัฒนาระบบการจัดสอบ การสร้างข้อสอบวิทยาศาสตร์ที่จะสร้างขึ้นจะวัดสมรรถนะขั้นสูงทางวิทยาศาสตร์จำนวน 3 ด้าน ตามกรอบการประเมินนักเรียนระดับนานาชาติ (PISA) คือ การระบุประเด็นทางวิทยาศาสตร์ (identifying scientific issues) สมรรถนะการอธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์ (Explaining phenomena scientifically) และสมรรถนะการใช้หลักฐานทางวิทยาศาสตร์ (using scientific evidence) ส่วนกรอบการพัฒนาระบบการจัดสอบ จะประกอบด้วยระบบการนำเข้าข้อสอบ ระบบการเขียน/ปรับข้อสอบ ระบบการสอบผ่านอินเทอร์เน็ต ระบบการวิเคราะห์ข้อสอบและรายงานผลการทดสอบ การนำเสนอการทดสอบจะผ่านอินเทอร์เน็ต ซึ่งนักเรียนทั่วไปประเทศสามารถเข้ามาทดสอบได้ กรอบการวิจัยทั้งสองกรอบอธิบายได้ดังแผนภาพต่อไปนี้



สมมติฐาน

จากผลการศึกษาเอกสารและงานวิจัยเกี่ยวกับการพัฒนาแบบทดสอบวิทยาศาสตร์ผ่านเทคโนโลยี เพื่อประเมินทักษะขั้นสูงของการเรียนรู้ พบว่า แบบทดสอบนี้สามารถวัดความรู้และทักษะขั้นสูงได้อย่างเหมาะสม (ตามหลักการวัดผลทางการศึกษาในด้านความตรงและความเที่ยง) เช่น งานวิจัยของ Quellmalz, Davenport, Timms, DeBoer, Jordan, Huang, Buckley (2013) ที่พบว่าการทดสอบผ่านเทคโนโลยีสามารถวัดความรู้และทักษะขั้นสูงของผู้เรียนได้ ดังนั้นผู้วิจัยจึงกำหนดสมมติฐานว่า แบบทดสอบที่จัดสอบผ่านระบบอินเทอร์เน็ตที่พัฒนาขึ้นสามารถวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นได้อย่างมีคุณภาพในด้านความตรงและความเที่ยง

ประโยชน์ที่คาดว่าจะได้รับ

1. ด้านวิชาการ

การวิจัยนี้เป็นการบูรณาการการทดสอบทางการศึกษาและเทคโนโลยีทางการศึกษาเข้าด้วยกัน ซึ่งถือเป็นแนวคิดใหม่ๆ ในการประเมินทางการศึกษาของประเทศไทย การวิจัยนี้จะพัฒนาการทดสอบการรู้วิทยาศาสตร์ด้วยรูปแบบที่ทันสมัย และมีความเป็นไปได้ในการนำไปใช้ ซึ่งอาจเป็นตัวอย่างเชิงวิชาการในการวัดและประเมินผลทางการศึกษาให้กับหน่วยงานทางการศึกษาต่างๆ เช่น สำนักทดสอบทางการศึกษาแห่งชาติ (สทศ.) หรือหน่วยงานทางการศึกษาที่จัดการทดสอบอื่นๆ นอกจากนี้ยังได้ผลลัพธ์เป็นเครื่องมือการประเมินวิทยาศาสตร์ที่สามารถวัดการรู้วิทยาศาสตร์ได้อย่างลึกซึ้ง

2. ด้านการนำไปปฏิบัติ

ผลลัพธ์ของโครงการอันหนึ่ง คือ การเผยแพร่การทดสอบการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ต ซึ่งเป็นเทคโนโลยีที่เข้าถึงประชาชนได้จำนวนมาก ดังนั้น ครูสามารถนำนักเรียนมาเข้า

รับการทดสอบ เพื่อนำผลการทดสอบไปปรับปรุงการเรียนการสอนตามหลักของการประเมินเพื่อการเรียนรู้ นอกจากนี้ ผลผลิตจากการวิจัยยังมีประโยชน์ในด้านการฝึกนักเรียนให้มีความพร้อมต่อการประเมิน PISA ในอนาคต ซึ่งในอนาคต PISA จะจัดสอบผ่านคอมพิวเตอร์เพื่อวัดความสามารถขั้นสูง ดังนั้นการที่นักเรียนมีความคุ้นชินกับการทดสอบผ่านเทคโนโลยีจะทำให้ไม่มีปัญหาเมื่อต้องสอบจริง ซึ่งจะทำให้การวัดนั้นสะท้อนคุณภาพการศึกษาของประเทศอย่างแท้จริง

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสำหรับงานวิจัยเรื่องนี้ มีประเด็นของการศึกษา 6 ตอนดังนี้

1. กรอบการประเมินสมรรถนะด้านวิทยาศาสตร์ของโครงการประเมินนักเรียนระดับนานาชาติ (PISA)

2. ระบบการประเมินเพื่อพัฒนาการเรียนรู้วิทยาศาสตร์: ระบบการประเมินที่สมดุล (balanced assessment)

3. การทดสอบผ่านคอมพิวเตอร์

4. การออกแบบการทดสอบวิทยาศาสตร์ด้วยเทคโนโลยีที่มีคุณภาพ

5. คุณภาพของการทดสอบ

6. งานวิจัยเกี่ยวกับปัญหาของการเรียนการสอนวิทยาศาสตร์

ซึ่งแต่ละตอนมีรายละเอียดดังนี้

1. กรอบการประเมินสมรรถนะด้านวิทยาศาสตร์ของโครงการประเมินนักเรียนระดับนานาชาติ (PISA)

PISA กำหนดว่าผู้เรียนที่มีความสามารถด้านวิทยาศาสตร์ควรมีลักษณะ 4 อย่าง คือ 1) มีความรู้ทางวิทยาศาสตร์ และใช้ความรู้วิทยาศาสตร์ในการระบุคำถาม หาความรู้ใหม่ อธิบายปรากฏการณ์ทางวิทยาศาสตร์ และหาข้อสรุปในประเด็นเกี่ยวกับวิทยาศาสตร์โดยมีหลักฐาน 2) เข้าใจลักษณะของวิทยาศาสตร์ว่าเป็นรูปแบบหนึ่งของความรู้ของมนุษย์และการหาความรู้ 3) มีความตระหนักว่าวิทยาศาสตร์และเทคโนโลยีทำให้เกิดวัสดุ ปัญญา และสิ่งแวดล้อมทางวัฒนธรรม และ 4) เข้าร่วมในประเด็นวิทยาศาสตร์ ด้วยการมีความคิดทางวิทยาศาสตร์ในฐานะของคนที่ไม่เชื่ออะไรง่ายๆ

PISA ประเมินสมรรถนะของนักเรียนอายุ 15 ปี ซึ่งเป็นนักเรียนระดับมัธยมศึกษาตอนต้น ซึ่งเป็นวัยที่ผ่านการศึกษาภาคบังคับของประเทศส่วนใหญ่ การออกแบบข้อสอบประเมินสมรรถนะวิทยาศาสตร์ของโครงการ PISA จะกำหนดโครงสร้างของคำถามไว้ 4 ประเด็น คือ 1) ประเภทคำตอบ 2) สมรรถนะ 3) ความรู้ 4) บริบทและขอบเขตของการนำไปใช้ ดังนี้

9.3.1 ประเภทคำตอบ

การประเมิน PISA ใช้ข้อสอบที่มีลักษณะหลากหลาย คือ ข้อสอบเลือกตอบ (Multiple-choice items (4-5 ตัวเลือก) ข้อสอบเลือกตอบ 2 ตัวเลือก เช่น ใช่ หรือ ไม่ใช่, เห็นด้วย หรือ ไม่เห็นด้วย ข้อสอบเลือกตอบแบบซับซ้อน (complex multiple-choice items) ตอบสั้น และตอบยาว

9.3.2 สมรรถนะ

PISA กำหนดสมรรถนะด้านการรู้วิทยาศาสตร์ ที่ต้องประเมินไว้ 3 สมรรถนะ คือ

1. ระบุประเด็นทางวิทยาศาสตร์ (identifying scientific issues) ประกอบด้วย การรู้ว่าประเด็นใดบ้างที่สามารถตรวจสอบได้อย่างวิทยาศาสตร์ การระบุคำสำคัญที่ต้องใช้ค้นหาข้อมูลทางวิทยาศาสตร์ และการรู้ลักษณะสำคัญของการตรวจสอบทางวิทยาศาสตร์

2. อธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์ (explaining phenomena scientifically) ประกอบด้วยการใช้ความรู้วิทยาศาสตร์ในสถานการณ์ที่กำหนด การอธิบายหรือแปลผลปรากฏการณ์อย่างวิทยาศาสตร์ และทำนายการเปลี่ยนแปลง การบอกถึงการบรรยาย อธิบาย และทำนายที่เหมาะสม

3. ใช้หลักฐานทางวิทยาศาสตร์ (Using scientific evidence) ประกอบด้วย การแปลผลหลักฐานทางวิทยาศาสตร์ สรุป และสื่อสารให้เข้าใจได้ การระบุข้อตกลงเบื้องต้น หลักฐาน และเหตุผลเบื้องหลังข้อสรุป และการสะท้อนนัยของการนำผลการพัฒนาทางวิทยาศาสตร์และเทคโนโลยีไปใช้ในสังคม

9.3.3 ความรู้

กรอบการประเมินการรู้วิทยาศาสตร์ของ PISA มีการประเมินความรู้ด้านวิทยาศาสตร์ (knowledge of science) และความรู้เกี่ยวกับวิทยาศาสตร์ (knowledge about science) ดังนี้

ความรู้ด้านวิทยาศาสตร์ เป็นความรู้ที่เกิดจากการเรียนวิทยาศาสตร์ ประกอบด้วย

- ระบบกายภาพ
- ระบบสิ่งมีชีวิต
- ระบบโลกและอวกาศ
- ระบบเทคโนโลยี

ความรู้เกี่ยวกับวิทยาศาสตร์ เป็นความรู้ที่ไม่จำเป็นต้องเกิดจากการเรียน แต่เป็นความรู้ที่นำมาใช้เมื่อต้องเกี่ยวข้องกับวิทยาศาสตร์ โดยมี 2 มิติที่ประเด็น คือ การหาความรู้ด้วยกระบวนการวิทยาศาสตร์ และการอธิบายทางวิทยาศาสตร์

9.3.4 บริบทและขอบเขตของการนำไปใช้

การประเมิน PISA ได้ออกแบบบริบทเกี่ยวกับการนำวิทยาศาสตร์ไปใช้ในบริบท 3 ด้าน คือ บริบทเกี่ยวกับตนเอง สังคม และโลก ในขอบเขตของการนำไปใช้ 5 ด้าน คือ

- สุขภาพ
- ทรัพยากรธรรมชาติ
- สิ่งแวดล้อม
- อันตราย
- วิทยาศาสตร์และเทคโนโลยีใหม่ๆ

2. ระบบการประเมินเพื่อพัฒนาการเรียนรู้วิทยาศาสตร์: ระบบการประเมินที่สมดุล (balanced assessment)

จากปัญหาของการเรียนการสอนวิทยาศาสตร์ที่กล่าวมา จะเห็นว่าการพัฒนาการเรียนการสอนวิทยาศาสตร์มีความจำเป็น การพัฒนานี้มีหลายองค์ประกอบที่จะช่วยกันยกระดับคุณภาพการ

เรียนการสอนวิทยาศาสตร์ ซึ่งส่วนใหญ่ประกอบด้วย การจัดการเรียนการสอน หลักสูตร มาตรฐาน และการประเมินผลการศึกษา ในงานวิจัยนี้มุ่งศึกษาองค์ประกอบของการประเมิน จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการประเมินที่ช่วยยกระดับคุณภาพการศึกษา พบว่า ระบบการประเมินที่นักวิชาการกล่าวถึงกันมากคือ ระบบการประเมินที่สมดุล (balanced assessment) ซึ่งเป็นระบบที่มีการบูรณาการการเรียนการสอนและการประเมินเข้าด้วยกัน ดังมีรายละเอียดดังนี้

1. ความสำคัญ และความหมายของระบบการประเมินที่สมดุล (Balanced Assessment)

ในอนาคต สังคมและเศรษฐกิจจะมีส่วนสำคัญในการปรับเปลี่ยนการจัดการศึกษาของโรงเรียนจากการเรียนการสอนแบบดั้งเดิมไปสู่การสอนที่เน้นให้นักเรียนทุกคนได้เรียนรู้ตามมาตรฐานการศึกษามากขึ้น เมื่อนักเรียนไปโรงเรียน โรงเรียนต้องรับผิดชอบการเรียนของนักเรียนทุกคน ไม่ควรเกิดกรณีว่ามีนักเรียนจำนวนหนึ่งผ่านเกณฑ์ของหลักสูตร และมีนักเรียนอีกจำนวนหนึ่งไม่ผ่านเกณฑ์ของหลักสูตร โรงเรียนต้องส่งเสริมให้นักเรียนทุกคนได้มีโอกาสได้เรียนรู้และผ่านมาตรฐานการเรียนรู้ ดังนั้นบทบาทของการประเมินทางการศึกษาในอนาคตจะต้องไม่ใช่แค่การตัดเกรดเท่านั้น แต่ต้องให้สารสนเทศแก่ผู้เกี่ยวข้องเพื่อใช้ส่งเสริมให้นักเรียนได้เรียนรู้จนผ่านมาตรฐานการเรียน ดังนั้นรูปแบบและวิธีการประเมินที่ครูและบุคลากรทางการศึกษาควรใช้ คือการประเมินที่สมดุล ซึ่งมีรายละเอียดดังนี้

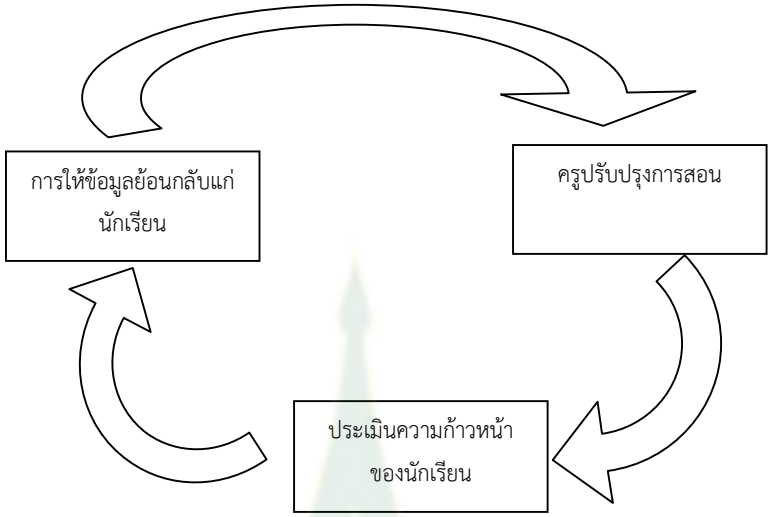
การประเมินทางการศึกษาที่สมดุล หมายถึง การประเมินที่ให้สารสนเทศเกี่ยวกับความรู้ความสามารถของผู้เรียนแก่ผู้ที่เกี่ยวข้องทุกฝ่าย เพื่อให้ผู้เกี่ยวข้องสามารถปรับปรุงการเรียนการสอน เพื่อนำไปสู่การพัฒนาการเรียนรู้ของผู้เรียน ผู้เกี่ยวข้องกับการจัดการศึกษาที่สำคัญ ได้แก่ นักเรียน ครู ผู้บริหารโรงเรียน ผู้ปกครอง และผู้บริหารเขตพื้นที่การศึกษา เป็นต้น ดังนั้น การประเมินทางการศึกษาต้องให้สารสนเทศที่ถูกต้องแก่บุคคลเหล่านี้

2. องค์ประกอบของการประเมินทางการศึกษาที่สมดุล

การประเมินทางการศึกษาที่สมดุลควรประกอบด้วยองค์ประกอบที่มีการประเมินทางการศึกษาที่ให้ข้อมูลเพื่อพัฒนาการศึกษาของผู้เรียนและครู ซึ่งควรประกอบด้วยวิธีการประเมินสามแบบ คือ การประเมินในชั้นเรียน (classroom assessment หรือ formative assessment) การประเมินเพื่อชี้วัด (interim assessment หรือ benchmark assessment) และการประเมินเพื่อสรุปรวม (summative assessment) ดังมีรายละเอียดต่อไปนี้

การประเมินในชั้นเรียน

การประเมินในชั้นเรียน หมายถึง การประเมินที่จัดทำโดยครูและนักเรียนในระหว่างการเรียนการสอน เพื่อให้ได้ข้อมูลสำหรับปรับปรุงการสอนของครู และปรับปรุงผู้เรียน การที่จะทำให้การประเมินประสบความสำเร็จต้องเกิดจากความร่วมมือของครูกับนักเรียน กล่าวคือ ภายหลังจากการประเมิน ครูต้องให้ข้อมูลย้อนกลับที่ชัดเจนแก่นักเรียน และต้องกระตุ้นให้นักเรียนรับผิดชอบการเรียนของตนเองโดยใช้วิธีการประเมินตนเองว่าทำได้ตามเป้าหมายของการเรียนรู้หรือไม่ ถ้ายังไม่สำเร็จ นักเรียนต้องร่วมมือกับครูเพื่อให้ทราบว่าต้องทำอะไรบ้างเพื่อให้สามารถแก้ไขข้อบกพร่องของนักเรียน ทั้งนี้ครูก็อาจต้องปรับเปลี่ยนการสอนจนกระทั่งนักเรียนสามารถเข้าใจหรือรอบรู้ตามวัตถุประสงค์ การกระทำนี้มีลักษณะเป็นวงจร (McMillan, 2008) ดังแสดงในรูปที่ 1



รูปที่ 1 วงจรการประเมินในชั้นเรียน

การใช้กระบวนการประเมินในชั้นเรียนและผลการประเมินเพื่อพัฒนาการเรียนรู้ของนักเรียน มีชื่อเรียกว่าการประเมินเพื่อการเรียนรู้ (assessment *for* learning) ผลการวิจัยพบว่า การประเมินเพื่อการเรียนรู้สามารถช่วยพัฒนาการเรียนรู้ของนักเรียนได้ดีกว่าวิธีการอื่น เช่น การลดขนาดชั้นเรียน (class size) หรือโครงการพัฒนาการศึกษา (intervention) นักเรียนที่ได้ประโยชน์มากที่สุดจากการประเมินเพื่อการเรียนรู้ คือ นักเรียนที่เรียนอ่อน ดังนั้นการประเมินเพื่อเรียนรู้จึงสามารถช่วยลดช่องว่างของการเรียนระหว่างนักเรียนเก่งกับนักเรียนอ่อนได้

การทำการประเมินเพื่อการเรียนรู้ให้ประสบความสำเร็จ ครูต้องมีความสามารถและปฏิบัติได้ดีในประเด็นต่อไปนี้ (Brookhart, 2008; Stiggins, 2008)

1. กำหนดจุดมุ่งหมายของการประเมินให้ชัดเจน ตลอดจนต้องทราบว่าข้อมูลที่ต้องการคืออะไร
2. กำหนดเป้าหมายของการเรียนรู้ให้ชัดเจน นักเรียนเข้าใจได้
3. ใช้วิธีการประเมินที่เหมาะสมกับเนื้อหา และหลากหลาย เช่น การประเมินตนเอง แฟ้มสะสมงาน และแบบทดสอบ
4. ให้ข้อมูลย้อนกลับแก่นักเรียนอย่างชัดเจน เข้าใจง่าย และ ทันเวลา เพื่อให้นักเรียนทราบว่าควรทำอะไร อย่างไร ต่อไป
5. สร้างแรงจูงใจให้นักเรียนไม่ย่อท้อกับการเรียน และตั้งใจทำให้ดียิ่งขึ้น
6. พัฒนาผู้เรียนให้รับผิดชอบผลการเรียนของตนเอง และประเมินตนเองตามความเป็นจริง

การประเมินเพื่อสรุปรวม

การประเมินเพื่อสรุปรวม เป็นการประเมินที่ใช้หลังการเรียนการสอน การประเมินแบบนี้ใช้เพื่อประเมินความรู้ของนักเรียนว่านักเรียนมีความรู้ความสามารถเพียงใด โดยทั่วไป คนส่วนใหญ่

เข้าใจว่าการประเมินเพื่อสรุปรวมใช้สำหรับการตัดเกรด หรือคัดเลือกผู้เรียน เท่านั้น ทว่า ปัจจุบันนี้ การประเมินเพื่อสรุปรวมได้ถูกนำไปใช้สำหรับการประเมินเพื่อประเมินคุณภาพการจัดการศึกษาของโรงเรียน หรือหน่วยงานที่รับผิดชอบการศึกษาอื่นๆ มากยิ่งขึ้น การประเมินเพื่อสรุปรวมที่ใช้ในการประเมินคุณภาพการศึกษาเป็นการประเมินมีลักษณะเป็นการประเมินการศึกษาในภาพรวมที่ดำเนินงานโดยบุคคลภายนอกโรงเรียน ผลการประเมินจะถูกนำไปใช้เพื่อกำหนดคุณภาพของบุคคลากรและหน่วยงานที่รับผิดชอบกับการจัดการศึกษา ไม่ใช่การประเมินนักเรียนรายบุคคล

เมื่อกล่าวถึงระบบการประเมินที่สมดุล การประเมินเพื่อสรุปรวมจึงมีสองส่วน คือ การประเมินโดยครูหลังการเรียนการสอนเพื่อการตัดเกรด และการประเมินเพื่อสรุปรวมที่ดำเนินโดยองค์กรภายนอกที่ทำขึ้นเพื่อประเมินคุณภาพการสอนของโรงเรียน ประโยชน์ของการประเมินทั้งสองแบบนี้มีความแตกต่างกัน กล่าวคือ แน่นอนว่า การประเมินหลังการสอนโดยครูก็เพื่อใช้ประเมินความรู้รวบยอดของนักเรียนของแต่ละหน่วยการเรียนรู้ หรือเพื่อการตัดเกรดหลังสิ้นสุดภาคการศึกษา ส่วนการประเมินโดยบุคคลภายนอกมีประโยชน์ในระดับนโยบายและการรับรองคุณภาพการศึกษา อย่างไรก็ตาม ในบริบทของการประเมินอย่างสมดุล โรงเรียนและครูควรนำผลการประเมินทั้งสองแบบมาวิเคราะห์จุดอ่อนและจุดแข็งของการจัดการศึกษา เพื่อให้ครูจัดการเรียนการสอนให้ดีขึ้นต่อไปในอนาคต การวิเคราะห์เปรียบเทียบผลการประเมินของนักเรียนที่สอนโดยครูคนละคนกัน ในโรงเรียนเดียวกัน จะให้ข้อมูลที่เป็นประโยชน์มากสำหรับการปรับปรุงการสอนในอนาคต เพราะการวิเคราะห์จะทำให้ทราบว่าผลการเรียนที่แตกต่างกันของนักเรียนเกิดจากการสอนของครูที่ต่างกัน หรือไม่ อย่างไร การสอนแบบใดดีที่สุด การวิเคราะห์เปรียบเทียบนี้อาจทำในรูปของคณะกรรมการครู ซึ่งจะช่วยให้ครูเกิดการเรียนรู้ และได้พัฒนาตนเองจากการเรียนรู้ร่วมกัน นอกจากนี้ โรงเรียนอาจนำผลการประเมินของโรงเรียนที่ประเมินโดยองค์กรภายนอกมาวิเคราะห์เปรียบเทียบกับโรงเรียนอื่น ก็จะทำให้ได้สารสนเทศอื่นเพื่อพัฒนาโรงเรียนต่อไปในอนาคต

การประเมินเพื่อชี้วัด

การประเมินเพื่อชี้วัด มีลักษณะของการประเมินที่อยู่ระหว่างการประเมินในชั้นเรียน และการประเมินเพื่อสรุปรวม กล่าวคือ ไม่ใช่การประเมินที่ดำเนินการโดยครูในระหว่างการสอน แต่ก็ไม่ใช่การประเมินหลังสิ้นสุดภาคการศึกษา หากแต่เป็นการประเมินในระหว่างที่มีการเรียนการสอนโดยบุคคลอื่นที่ไม่ใช่ครู และเป็นการประเมินการศึกษาในภาพรวม ไม่ใช่การประเมินนักเรียนเป็นรายบุคคล โดยส่วนใหญ่ ผู้ที่ดำเนินการประเมินเพื่อชี้วัด คือ หน่วยงานภายนอกที่รับผิดชอบการจัดการศึกษา เช่น เขตพื้นที่การศึกษา หรือ หน่วยงานจากส่วนกลาง เป็นต้น นอกจากนี้ การประเมินเพื่อชี้วัดจะมีการประเมินเป็นระยะๆ อาจจะทุกเดือน หรือทุกสองเดือนในระหว่างที่มีการเรียนการสอน ซึ่งต่างจากการประเมินเพื่อสรุปรวมที่มีการประเมินครั้งเดียว

การประเมินเพื่อชี้วัดอาจมีวัตถุประสงค์อย่างใดอย่างหนึ่ง หรือหลายอย่างประกอบกันดังต่อไปนี้ (Pirie, Marion, & Gong, 2007)

1. เพื่อปรับปรุงการสอน

วัตถุประสงค์ข้อนี้เหมือนกับการประเมินในชั้นเรียนมาก คือการประเมินเพื่อปรับเปลี่ยนการสอนให้สอดคล้องกับความต้องการของนักเรียน การประเมินเพื่อชี้วัดที่ตอบสนองวัตถุประสงค์ข้อนี้

คือ การดำเนินการประเมินว่าตรวจสอบว่านักเรียนส่วนใหญ่มีจุดอ่อนด้านใดบ้างที่ต้องการแก้ไขอย่างเร่งด่วน ณ จุดใดจุดหนึ่งระหว่างที่มีการเรียนการสอน หลังการประเมิน ครูและโรงเรียนจะได้รับผลการประเมินเป็นภาพรวมของโรงเรียนหรือห้องเรียน จากนั้นการดำเนินการปรับปรุงแก้ไขการสอนมักจะขึ้นอยู่การตัดสินใจของครูและโรงเรียน

2. เพื่อประเมินหลักสูตรโปรแกรมการศึกษา หรือวิธีการสอน

การใช้การประเมินเพื่อชี้วัดคุณภาพของโปรแกรมการศึกษา หรือวิธีการสอนมักเป็นการประเมินขนาดใหญ่ที่มีโรงเรียนเข้าร่วมจำนวนมาก เพื่อประเมินว่าโปรแกรมการจัดการศึกษาแบบใดหรือวิธีการสอนแบบใดมีคุณภาพ ผู้ที่ได้รับประโยชน์จากการประเมินในลักษณะนี้ไม่จำเป็นต้องเป็นนักเรียนที่ถูกประเมิน ณ เวลานั้น แต่อาจเป็นนักเรียนในอนาคตก็ได้ เพราะผลการประเมินจะนำไปใช้เพื่อปรับปรุงหลักสูตรและวิธีการสอนของครูให้ดีขึ้นในอนาคต

3. เพื่อทำนายผลสัมฤทธิ์ของนักเรียน

การประเมินเพื่อทำนายผลสัมฤทธิ์ทางการเรียนของนักเรียนในระหว่างที่มีการเรียนการสอนใช้เมื่อต้องการทำนายว่า การเรียนการสอน ณ ปัจจุบันจะสามารถช่วยให้นักเรียนสอบผ่านในการประเมินที่จะมีขึ้นในอนาคตหรือไม่ เช่น การประเมินเพื่อสำเร็จการศึกษา และการประเมินของหน่วยงานภายนอก การประเมินแบบนี้จะช่วยให้ครู และโรงเรียนทราบว่านักเรียนมีความรู้ความสามารถใกล้หรือไกลจากเป้าหมายของการประเมินที่นักเรียนต้องถูกประเมินในอนาคต ซึ่งดีกว่าการปล่อยให้ครูสอนไปตามเนื้อหาโดยไม่สนว่านักเรียนจะประสบความสำเร็จหรือไม่ การประเมินแบบนี้เริ่มมีบทบาทมากในยุคปัจจุบันสำหรับการรับรองและตรวจสอบคุณภาพการศึกษา เพราะผลการประเมินจะทำให้โรงเรียนทราบว่านักเรียนต้องได้รับการพัฒนาในเนื้อหาสาระใดบ้างก่อนที่จะเข้าสู่การประเมินจากภายนอก

9.2.2 การออกแบบการประเมินที่สมดุลอย่างมีคุณภาพ

จากที่กล่าวมาข้างต้น ว่า การประเมินที่สมดุลต้องให้ข้อมูลสำหรับการปรับปรุงการศึกษาแก่ผู้ที่มีส่วนเกี่ยวข้องทุกฝ่ายได้ใช้ร่วมกัน การใช้ผลการประเมินในชั้นเรียน ผลการประเมินเพื่อสรุปรวม และผลการประเมินเพื่อชี้วัดของผู้เกี่ยวข้องทั้งหลาย จะเอื้อให้มีข้อมูลที่หลากหลายสำหรับการพัฒนาคุณภาพการศึกษา การออกแบบการประเมินที่สมดุลจะมีคุณภาพเพียงใด ขึ้นอยู่กับประเด็นที่สำคัญต่อไปนี้

1. การประเมินทั้งสามแบบต้องประเมินเป้าหมายหรือมาตรฐานเดียวกัน
2. ครูต้องจัดการเรียนการสอนตามมาตรฐาน
3. เครื่องมือประเมินต้องวัดเนื้อหาที่กำหนดในมาตรฐาน และสอดคล้องกับวัตถุประสงค์ของการประเมินและการเรียนการสอนของครู
4. เครื่องมือประเมินมีคุณภาพตามมาตรฐานของการวัดและการทดสอบทางการศึกษา
5. นักเรียนมีโอกาสได้เรียนรู้สาระที่กำหนดในมาตรฐาน
6. มีการรายงานข้อมูลหรือผลการประเมินที่เข้าใจง่าย และทันเวลา

3. การทดสอบผ่านคอมพิวเตอร์

การทดสอบผ่านคอมพิวเตอร์

ประเภทของการทดสอบผ่านคอมพิวเตอร์

4. การออกแบบการทดสอบวิทยาศาสตร์ด้วยเทคโนโลยีที่มีคุณภาพ

การพัฒนาข้อสอบให้เป็นข้อสอบที่วัดความรู้และสมรรถนะของผู้เรียนผ่านสถานการณ์จำลองที่นำเสนอด้วยเทคโนโลยีเป็นความก้าวหน้าทางการทดสอบ ซึ่งเกิดจากการบูรณาการความรู้ด้านการทดสอบ และการประเมินเข้ากับเทคโนโลยีที่ทันสมัย Quellmalz, Davenport, Timms, DeBoer, Jordan, Huang, Buckley (2013) เสนอแนะว่าการออกแบบการทดสอบวิทยาศาสตร์ด้วยเทคโนโลยีที่มีประสิทธิภาพ มีควรดำเนินการดังนี้

1. เลือกบริบทที่เหมาะสมกับวิทยาศาสตร์

โครงการประเมินการศึกษาด้านวิทยาศาสตร์ระดับนานาชาติในปัจจุบัน เช่น Trends in Mathematics and Science Study (TIMSS), Programme for International Student Assessment (PISA), และ National Assessment of Educational Progress (NAEP) มีการประเมินความสามารถในการแก้ปัญหา และเหตุผลในการดำเนินการแก้ปัญหาของผู้สอบ ซึ่งสอดคล้องกับกรอบการจัดการศึกษาวิทยาศาสตร์ และมาตรฐานวิทยาศาสตร์ในอนาคต ที่เสนอแนะว่าการประเมินวิทยาศาสตร์ควรใช้หลักการออกแบบการวิจัย และข้อเสนอแนะเหล่านี้ยังสอดคล้องกับแนวคิดของนักวิชาการด้านสมอง (cognitive science) หากเราพิจารณาเหตุการณ์ที่เกิดขึ้นในโลกความจริง จะเห็นว่า คนที่เก่งและเป็นผู้เชี่ยวชาญมักจะมีโครงสร้างของความรู้ขนาดใหญ่ มีการจัดอย่างเป็นระบบ และความรู้ที่จัดเก็บเชื่อมต่อประสานงานกันอย่างดี รวมทั้งมีทักษะการแก้ปัญหาดีมาก ดังนั้น เพื่อให้นักเรียนมีสมรรถนะเหล่านี้ การสร้างบรรยากาศของการเรียนรู้ที่ต้องการให้นักเรียนได้แสดงออกถึงความรู้ที่มีอยู่ในตัว จะต้องออกแบบบริบทของการเรียนอย่างดี ซึ่งเป็นไปตามทฤษฎีการเรียนรู้ (learning theory) เช่นเดียวกันในเรื่องของการทดสอบ บริบทของการทดสอบที่เหมาะสมควรเป็นบริบทที่ให้ผู้สอบได้ใช้ความรู้ที่หลากหลายอย่างบูรณาการ ใช้ทักษะการแก้ปัญหาเชิงกลยุทธ์ และได้ความรู้จากการทดสอบด้วย (Linn & Eylon, 2011; Quellmazl & Pellegrino, 2009)

2. ออกแบบการประเมินและวัตถุประสงค์การเรียนรู้ให้สอดคล้องกัน

หากนักเรียนมีความเข้าใจวิทยาศาสตร์อย่างลึกซึ้ง นักเรียนต้องเข้าใจคอนเซ็ปที่สำคัญของวิทยาศาสตร์ และสามารถที่จะใช้ความรู้เพื่อทำนาย และดำเนินการตรวจสอบทางวิทยาศาสตร์ได้ ดังนั้นการพัฒนาการทดสอบวิทยาศาสตร์จึงต้องวัดให้ถึงความสามารถเชิงกลยุทธ์ที่จำเป็นต่อการเข้าใจและอธิบายระบบวิทยาศาสตร์ที่ซับซ้อน รวมถึงความสามารถในการตรวจสอบทางวิทยาศาสตร์ของผู้สอบ

นักวิชาการเสนอว่า การออกแบบการประเมินวิทยาศาสตร์ควรออกแบบให้การประเมินนั้นสอดคล้องกับวัตถุประสงค์ของการเรียนรู้ หรือวัตถุประสงค์ของการสอน โดย Mislevy,

Almond, และ Lukas (2003) เสนอว่าการออกแบบการประเมินนี้มีสามส่วนสำคัญที่ต้องดำเนินการ ได้แก่

2.1 การออกแบบโมเดลนักเรียน (student model) หมายถึงการระบุวัตถุประสงค์การเรียนรู้ให้ชัดเจน ซึ่งส่วนใหญ่นำมาจากมาตรฐานในหลักสูตร

2.2 การออกแบบงานที่ทำ (task model) ซึ่งเป็นการอธิบายลักษะของข้อสอบ และบริบท หรือสิ่งแวดล้อมที่ต้องการจัดไว้ในข้อสอบ เพื่อให้นักเรียนได้แสดงความรู้ความสามารถ จนผู้ประเมินสามารถตัดสินใจได้ว่านักเรียนมีความสามารถตามที่ต้องการวัดหรือไม่ การออกแบบงานที่ทำในส่วนนี้เกี่ยวข้องกับการจำลองสถานการณ์โดยใช้เทคโนโลยี

2.3 การออกแบบหลักฐาน (evidence model) หมายถึงการอธิบายผลการตอบของนักเรียนภายใต้การจำลองสถานการณ์ ที่ถือว่าเป็นหลักฐานของความรู้และสมรรถนะที่ต้องการวัด หรือเป็นหลักฐานว่านักเรียนมีความรู้ความสามารถที่ยอมรับได้

3. การลดความคลาดเคลื่อนจากการสอบ

การใช้เทคโนโลยีในการทดสอบอาจทำให้มีความคลาดเคลื่อนในการวัด เพราะว่าการออกแบบข้อสอบและการทดสอบผ่านเทคโนโลยี จะดำเนินการผ่านหน้าจอคอมพิวเตอร์ รูปแบบตัวอักษร สีสัน และการนำเสนอผ่านหน้าจอที่แปลกใหม่ หรือเป็นการทดสอบที่ให้ผู้สอบคลิ๊กเมาส์เพื่อดูข้อมูลประกอบการตอบ สิ่งเหล่านี้อาจมีผลต่อความรู้สึก และผู้สอบต้องใช้สมองในการดำเนินงานมาก (cognitive demand) ดังนั้นผู้ออกแบบการทดสอบด้วยเทคโนโลยีควรควบคุมให้มีความคลาดเคลื่อนในการทดสอบให้น้อยที่สุด นอกจากนี้ การนำเสนอข้อสอบผ่านสถานการณ์จำลองหรือแอนิเมชั่นอาจช่วยเพิ่มความสนใจของผู้สอบ และกระตุ้นให้ผู้สอบคิดหรือใช้สมองมากกว่าการทดสอบแบบปกติ ผลการวิจัยเสนอแนะว่าเพื่อลดภาระของการใช้สมองของผู้สอบ ควรดำเนินการดังนี้

1) ออกแบบงานที่ทำและรูปแบบของงานให้สอดคล้องกัน ตัวอย่างเช่น การใช้แอนิเมชั่นควรใช้เมื่อจำเป็นเท่านั้น หรือใช้เมื่อต้องการให้ผู้สอบสังเกตการเปลี่ยนแปลงของปรากฏการณ์ที่สังเกตได้ยากในสถานการณ์จริง

2) ให้ผู้สอบควบคุมการสอบได้ เช่น การหยุด หรือ การดูซ้ำ เป็นต้น และ

3) มีการแจ้งเตือนหากมีการเปลี่ยนแปลง

5. คุณภาพของการทดสอบ

มาตรฐานด้านการวัดทางจิตวิทยาฉบับแรกที่เผยแพร่ในปี ค.ศ. 1954 มีการเขียนบทอธิบายเกี่ยวกับความตรง (validity) ซึ่งเป็นประเด็นสำคัญด้านการวัดในยุคนั้น และมาตรฐานฉบับนี้ได้อธิบายว่า ความตรงเป็นสิ่งที่บ่งบอกระดับของความสำเร็จตามวัตถุประสงค์ของการวัดของแบบวัด ซึ่งการอธิบายในมาตรฐานนี้เน้นให้เห็นว่าความตรงมีระดับของความมากน้อย ไม่ใช่เป็นการแบ่งประเภทที่แยกออกจากกันอย่างชัดเจน คือ “ตรง” หรือ “ไม่ตรง” และมาตรฐานด้านการวัดฉบับนี้ได้กำหนดว่าความตรงมีทั้งหมดมี 4 ประเภท คือ ความตรงเชิงเนื้อหา (content validity) ความตรงเชิงทำนาย (predictive validity) ความตรงตามสภาพ (concurrent validity) และความตรงเชิงโครงสร้าง (construct validity) สำหรับการทำมาตรฐานด้านการวัดฉบับแรกนี้มีนักวัดผลหลายคนร่วมกันเขียน

โดยมีปรมาจารย์ด้านการวัดผลในยุคนั้นจำนวนสองคน คือ มีห์ล (Meehl) และครอนบาค (Cronbach) ทำหน้าที่เป็นบรรณาธิการ

ความตรงที่กำหนดในมาตรฐาน ประกอบด้วย ความตรงเชิงเนื้อหา ความตรงเชิงทำนาย ความตรงตามสภาพ และความตรงเชิงโครงสร้าง มีความหมายและลักษณะดังนี้

1. ความตรงเชิงเนื้อหา หมายถึง การที่เครื่องมือวัดสามารถวัดได้ครอบคลุมเนื้อหาหรือสาระสำคัญในสิ่งที่ต้องการวัด

2. ความตรงตามสภาพหรือความตรงร่วมสมัย หมายถึง ความสามารถของเครื่องมือวัดที่วัดได้ตรงหรือสอดคล้องกับผลการวัดอื่นๆ ในขณะนั้น

3. ความตรงเชิงพยากรณ์ หมายถึง ความสามารถของเครื่องมือที่สามารถทำนายความสามารถของบุคคลในอนาคต

4. ความตรงเชิงโครงสร้าง หมายถึง ความสามารถของเครื่องมือที่จะวัดคุณลักษณะหรือพฤติกรรมตามโครงสร้างของทฤษฎี

ในปี ค.ศ.1966 ได้มีการปรับปรุงมาตรฐานการทดสอบขึ้น โดยการให้นิยามของความตรงก็ยังคงไม่ต่างจากนิยามในมาตรฐาน ปี ค.ศ. 1954 แต่มีประเด็นสำคัญอย่างหนึ่งที่เปลี่ยนแปลงไปจากเดิม คือ มีการจัดกลุ่มของประเภทของความตรงเสียใหม่ โดยลดจาก 4 ประเภท เหลือ 3 ประเภท ได้แก่ ความตรงเชิงเนื้อหา (content validity) ความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) และความตรงเชิงโครงสร้าง (construct validity) โดยความตรงตามเกณฑ์สัมพันธ์เป็นความสามารถของเครื่องมือวัดที่วัดได้ตรงกับเกณฑ์ภายนอก ซึ่งอาจเป็นเกณฑ์ในปัจจุบันหรือเกณฑ์ในอนาคตก็ได้ ความตรงตามเกณฑ์สัมพันธ์ แบ่งได้เป็น 2 ชนิด คือ ความตรงตามสภาพ (concurrent-related validity) และความตรงเชิงทำนาย (predictive validity)

ในปี ค.ศ. 1971 ครอนบาค เสนอว่า การตรวจสอบความตรงเป็นกระบวนการตรวจสอบความเหมาะสมของการทำนายหรือการอ้างถึงความสามารถของบุคคลจากผลการทดสอบที่บุคคลนั้นสอบ ครอนบาคพยายามเสนอแนวคิดเพื่อชี้ให้เห็นว่าการตรวจสอบความตรงควรทำในลักษณะเหมือนกับ***การประเมินแบบองค์รวม หรือบูรณาการ*** ไม่ใช่การแยกพิจารณาเป็นส่วนๆ เช่น การตรวจสอบความตรงเชิงเนื้อหา ความตรงตามเกณฑ์สัมพันธ์ และความตรงเชิงโครงสร้าง จากข้อเสนอในครั้งนี้ มีสิ่งที่สำคัญประการหนึ่งที่สามารถมองเห็นได้จากมุมมองของครอนบาคคือ ในฐานะที่ครอนบาคเป็นนักประเมิน จึงมองว่าการตรวจสอบความตรงเป็นกระบวนการคล้ายกับการประเมิน กล่าวคือ ***การตรวจสอบความตรงเป็นการประเมินของผู้พัฒนาแบบทดสอบในเชิงบูรณาการ*** ความคิดนี้มีบทบาทสำคัญต่อการให้นิยามความตรงในระยะต่อมาอย่างมาก

ครอนบาคได้เสนอความเห็นที่ท้าทายความคิดเห็นของนักวัดรุ่นก่อนๆ โดยตั้งคำถามว่า ในการตรวจสอบความตรง เรากำลังตรวจสอบอะไรกันแน่ และคำว่า “ตรง” คือความตรงของอะไร สุดท้ายครอนบาค (1971) จึงเสนอว่า สิ่งที่ควรประเมินจะมีความตรงในระดับมากหรือน้อยนั้น จริงๆ แล้วคือ **“ความหมายของคะแนน (meaning)” “การแปลผลคะแนน (interpretation)” และ “การนำคะแนนไปใช้(implication)”** จากข้อเสนอของครอนบาคตรงนี้มีนักวัดและนักวิชาการด้านการทดสอบหลายคนเห็นคล้อยตามกันในประเด็นที่ว่า ความตรงไม่ใช่เรื่องความตรงของแบบวัดหรือ

เครื่องมือวัด แต่เป็นความตรงของ “ความหมายของคะแนน” “การแปลผลคะแนน” และ “การนำคะแนนไปใช้” ตามที่กล่าวมาแล้ว

แนวคิดนี้นำไปสู่การให้นิยามของความตรงในยุคต่อมา เช่น เมสซิค (Messick, 1995) กล่าวว่า การตรวจสอบความตรง เป็น***การพิจารณาในเชิงประเมินแบบภาพรวม หรือองค์รวมว่าหลักฐานที่เก็บรวบรวมมาจะใช้ประเมินความตรงได้เหมาะสมหรือไม่*** รวมถึงการหาเหตุผลเชิงทฤษฎีที่สนับสนุนถึง***ความเหมาะสมและเพียงพอของการแปลความหมายและการใช้คะแนนการวัด/ประเมิน*** ได้มากน้อยเพียงใด ดังคำว่า

> “*validity is an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretation and actions on the basis of test scores or other modes of assessments” (Messick, 1995, p. 741*)

ในการอธิบายความหมายของความตรงให้เข้าใจตรงกันนั้น สำหรับชาวต่างประเทศที่ใช้ภาษาอังกฤษมักจะไม่มีปัญหา เพราะใช้คำว่า validity ซึ่งเป็นคำที่คุ้นเคยกันสำหรับผู้ใช้ภาษาอังกฤษ แต่สำหรับคนไทยเมื่อพูดถึงความตรง มักจะสื่อความหมายได้ไม่ครอบครอบคลุมประเด็นที่ครอนบาคเสนอ และถ้ากล่าวว่า “ตรง” ตามแนวคิดของครอนบาคที่กล่าวมาแล้ว ก็จะต้องกล่าวว่า ความตรงของ “ความหมายของคะแนน” “การแปลผลคะแนน” และ “การนำคะแนนไปใช้” ซึ่งคำเหล่านี้เป็นคำที่ค่อนข้างแปลกสำหรับคนไทยและจะทำให้สับสน หรือเข้าใจเรื่องนี้ไม่ชัดเจน ดังนั้นการที่จะอธิบายว่า “ตรง” อย่างไรนั้น เราต้องทำความเข้าใจนิยามของความตรงที่เสนอโดยเมสซิคอีกครั้งหนึ่ง ซึ่งหากจะสรุปอย่างสั้นๆ ตามความหมายของความตรงที่เสนอโดยเมสซิค ที่เสนอว่า ***การตรวจสอบความตรง ก็คือ การหาหลักฐานเพื่อนำมาประเมินว่าเราจะแปลความหมายคะแนน และใช้ผลคะแนนเพื่อวัตถุประสงค์ใดวัตถุประสงค์หนึ่งได้เหมาะสมระดับใด และหลักฐานนั้นเพียงพอที่จะสรุปผลการประเมินได้เหมาะสมเพียงใด*** ถ้าหลักฐานและเหตุผลที่นำมาอ้างยังไม่ดีหรือไม่เพียงพอ ก็อาจจะสรุปว่าการแปลผลและการใช้ผลคะแนนครั้งนั้นยังไม่เหมาะสม หรือยังไม่ควรนำไปใช้ ดังนั้น เพื่อให้เข้าใจมากขึ้น เวลาที่เรากล่าวถึงความตรง ความตรงจึงเป็นเรื่องของการประเมินความเหมาะสมของการแปลผล และการใช้คะแนน หากเข้าใจความตรงตามความหมายนี้จะช่วยให้เราเข้าใจว่า ***ความตรงไม่ใช่คุณสมบัติของแบบทดสอบหรือเครื่องมือวัด แต่เป็นคุณสมบัติของการแปลความหมาย และการใช้คะแนน*** เพราะการจะเกิดความเหมาะสมนั้นบางครั้งขึ้นอยู่กับการใช้ของผู้ใช้ ไม่ใช่อยู่ที่แบบทดสอบและกระบวนการตรวจสอบความตรงเป็นกระบวนการหาหลักฐานเพื่อใช้ในการประเมินความเหมาะสมของการใช้และการตีความคะแนนจากเครื่องมือวัด

มาตรฐานการวัดและการทดสอบทางจิตวิทยาฉบับที่ 5 เสนอว่า หลักฐานที่เราสามารถเก็บรวบรวมมาใช้ตรวจสอบความตรง มี 5 ประเภท ได้แก่

1) เนื้อหา (test content) การทดสอบต้องสามารถวัดเนื้อหาที่ต้องการทดสอบอย่างถูกต้อง และหากเป็นแบบทดสอบสองฉบับที่คู่ขนานกัน ควรเป็นแบบทดสอบที่วัดเนื้อหาเดียวกัน

2) กระบวนการตอบ (response process) กระบวนการหาคำตอบ และการได้คะแนนของผู้สอบต้องเป็นกระบวนการที่ถูกต้อง ประเด็นของการตรวจสอบ คือ การตรวจคำตอบว่าถูกต้องหรือไม่ และการวิเคราะห์ค่าคะแนนของผู้สอบ ซึ่งต้องสอดคล้องกับทฤษฎี หรือมีเหตุผลที่ยอมรับได้

3) โครงสร้างภายใน (internal structure) ความสัมพันธ์ภายในของสิ่งที่ต้องการวัด เช่น องค์ประกอบ ความเที่ยง และการทำหน้าที่ต่างกันของข้อสอบ (differential item functioning, DIF) ข้อสอบที่ทำหน้าที่ต่างกัน คือ ข้อสอบที่กลุ่มผู้สอบต่างกันกันที่มีความสามารถเท่าเทียมกัน มีโอกาสในการตอบถูกต่างกัน ข้อสอบที่มีลักษณะเช่นนี้ ถือว่าเป็นข้อสอบที่ไม่เป็นธรรม หรือลำเอียง การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสามารถบอกความเท่าเทียมของการทดสอบของกลุ่มผู้สอบหลายๆ กลุ่มที่มีภูมิหลังต่างกันได้ ซึ่งช่วยให้สามารถแปลความหมายคะแนนและเปรียบเทียบคะแนนระหว่างผู้สอบได้

4) ความสัมพันธ์กับตัวแปรอื่น (relation to other variables) คะแนนที่ได้จากการวัดต้องมีความสัมพันธ์กับตัวแปรอื่นตามทฤษฎี ซึ่งเป็นการตรวจสอบความหมายของคะแนนที่ได้จากการวัดว่ามีความเหมาะสมมากเพียงใด

5) ผลที่เกิดตามมา (consequences of testing) การใช้คะแนนสอบเพื่อวัตถุประสงค์ใดประสงค์หนึ่ง ไม่ควรเกิดผลเสียต่อผู้อื่น อันเนื่องมาจากความผิดพลาดของแบบทดสอบ ดังนั้นข้อสอบ และกระบวนการที่เกี่ยวข้องจึงต้องมีความถูกต้อง เพื่อป้องกันผลกระทบต่อผู้ที่มีส่วนได้เสีย

การตรวจสอบความตรง (validation) จึงเป็นกระบวนการที่ใช้เก็บรวบรวมข้อมูลเพื่อสนับสนุนการอ้างอิงถึงความสามารถหรือพฤติกรรมของผู้สอบโดยอ้างอิงจากผลการวัดหรือผลการทดสอบ การตรวจสอบความตรงเชิงเนื้อหาจึงมีลักษณะเป็นการรวบรวมหลักฐานเพื่อนำมาประเมินว่าการใช้และการแปลผลคะแนนจากการทดสอบมีความเหมาะสมระดับใด การตรวจสอบความตรงอาจต้องดำเนินการหลายวิธีเพื่อให้มั่นใจว่าการใช้และการแปลผลคะแนนจะมีความเหมาะสม

จากการศึกษากระบวนการตรวจสอบความตรงข้างต้น จะพบว่าเป็นการรวบรวมหลักฐานเกี่ยวกับการวัดและการทดสอบเพื่อใช้ประเมินความเหมาะสมของการใช้และการแปลความหมายคะแนน งานวิจัยเรื่องนี้จึงใช้หลักฐานหลายอย่างในการตรวจสอบความตรงหรือความเหมาะสมของคะแนนและการใช้คะแนน เช่น หลักฐานจากผลการวิเคราะห์ข้อสอบด้วยทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบข้อสอบ การวิเคราะห์องค์ประกอบเชิงยืนยัน การวิเคราะห์ความสัมพันธ์ของคะแนนเมตาคอกนิชันกับคะแนนของตัวแปรอื่น

6. งานวิจัยเกี่ยวกับปัญหาของการเรียนการสอนวิทยาศาสตร์

การวิจัยในประเทศไทยที่มีการศึกษาปัญหา และแนวทางการพัฒนาการเรียนการสอนวิทยาศาสตร์ประกอบด้วยการวิจัยเชิงทดลองและการวิจัยเชิงสำรวจ การวิจัยเชิงทดลองส่วนมากใช้ในการศึกษาเปรียบเทียบวิธีการสอนวิทยาศาสตร์ที่พัฒนาขึ้นใหม่กับวิธีการสอนแบบเก่า รูปแบบวิธีการสอนที่พัฒนาขึ้นส่วนมากเป็นรูปแบบของการสอนแบบสืบเสาะที่ให้ผู้เรียนหาความรู้ด้วยตนเอง ผลการศึกษาพบว่าวิธีการสอนที่พัฒนาขึ้นใหม่มีประสิทธิภาพในการเพิ่มความรู้ของนักเรียน

และมีประสิทธิภาพมากกว่าวิธีการสอนแบบเก่า ส่วนการวิจัยเชิงสำรวจส่วนมากใช้ในการศึกษาสภาพปัญหาของการเรียนการสอนวิทยาศาสตร์และความต้องการของครู และปัจจัยที่สัมพันธ์กับความสามารถทางวิทยาศาสตร์ของนักเรียน ผลการสำรวจส่วนใหญ่ พบว่าครูมีปัญหาด้านวิธีการสอน ขาดสื่อการสอน ขาดทรัพยากรทางการศึกษา และต้องการพัฒนาตนเองเพื่อให้มีความรู้มากขึ้น นักเรียนที่มีความสามารถทางวิทยาศาสตร์สูงตั้งใจเรียนและมีความรับผิดชอบมากกว่านักเรียนที่มีความสามารถทางวิทยาศาสตร์ต่ำ ตัวอย่างของงานวิจัยเหล่านี้มีรายละเอียดดังต่อไปนี้

สังวรณ์ งัดกระโทก (2552) ศึกษาปัญหาการสอนวิทยาศาสตร์ของไทยโดยใช้ข้อมูลผลการประเมินนักเรียนระดับนานาชาติ (PISA) ปี พ.ศ. 2549 เพื่อให้ได้สารสนเทศที่เป็นประโยชน์สำหรับการปฏิรูปการเรียนการสอนวิทยาศาสตร์ของไทย วัตถุประสงค์ย่อยของการวิจัยประกอบด้วย 1) เพื่อศึกษาความสัมพันธ์ของการสอนวิทยาศาสตร์ของครูและความสามารถด้านวิทยาศาสตร์ของนักเรียนพร้อมกับตรวจสอบว่าความสัมพันธ์เหล่านั้นได้รับผลกระทบจากการประกันคุณภาพการศึกษา เทคโนโลยีการศึกษา ข้อจำกัดด้านทรัพยากรทางการศึกษา และบริบทของโรงเรียนหรือไม่ 2) ตรวจสอบว่านักเรียนที่มีความสามารถสูงและนักเรียนที่มีคะแนนต่ำกว่าค่าเฉลี่ยนานาชาติมีความแตกต่างกันในปัจจัยใด และ 3) เพื่อศึกษาว่าปัจจัยใดบ้างที่สามารถทำนายความน่าจะเป็นในการได้คะแนนวิทยาศาสตร์สูงกว่าค่าเฉลี่ยนานาชาติ ในรายงานการวิจัย มีการอภิปรายผลการวิจัย และข้อเสนอแนะเชิงนโยบาย ผลการวิจัยที่สำคัญสรุปได้ดังนี้

1. ความสัมพันธ์ระหว่างการเรียนการสอนแบบสืบเสาะหาความรู้มีความสัมพันธ์ทางลบกับความสามารถด้านวิทยาศาสตร์ของนักเรียนไทย ผลการวิเคราะห์นี้ชี้ว่าวิธีการเรียนการสอนแบบสืบเสาะหาความรู้ที่ครูใช้มีคุณภาพต่ำ ปัญหาที่สำคัญของการสอนแบบนี้คือวิธีการสอนขาดคุณภาพ โดยด้านที่เป็นปัญหาได้แก่ การสอนแบบมีปฏิบัติการ การให้นักเรียนศึกษาทดลองด้วยตนเอง และปฏิสัมพันธ์ระหว่างครูและนักเรียน นอกจากนี้ผลการวิจัยยังพบว่า ในบริบทของสังคมไทย วิธีการสอนที่เน้นครูเป็นศูนย์กลางมีประสิทธิภาพมากกว่าการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลาง
2. ข้อจำกัดด้านทรัพยากรทางการศึกษาที่กระทบต่อคุณภาพการสอนวิทยาศาสตร์ประกอบด้วยการขาดวัสดุอุปกรณ์ทางวิทยาศาสตร์ การขาดบุคลากรสนับสนุน และการขาดครูที่มีความรู้
3. คุณภาพการสอนของครูในด้านการให้นักเรียนปฏิบัติการทดลองมีคุณภาพต่ำ และมีคุณภาพลดลงอีกเมื่อโรงเรียนขาดบุคลากรสนับสนุน
4. คุณภาพการด้านการสื่อสารและปฏิสัมพันธ์ของครูกับนักเรียนและคุณภาพของการให้นักเรียนศึกษาและทดลองด้วยตนเองไม่มีคุณภาพ และคุณภาพของการสอนทั้งสองวิธีนี้ลดลงไปอีกเมื่อครูขาดวัสดุและอุปกรณ์ประกอบการสอน
5. จำนวนคอมพิวเตอร์ที่เชื่อมต่ออินเทอร์เนตมีความสัมพันธ์ทางบวกอย่างมีนัยสำคัญกับความสามารถด้านวิทยาศาสตร์ของนักเรียน แต่ชั้นเรียนขนาดเล็กมีอิทธิพลทางลบต่อความสามารถด้านวิทยาศาสตร์ของนักเรียน
6. การดำเนินงานเกี่ยวกับการประกันคุณภาพการศึกษาของโรงเรียน และบรรยากาศการส่งเสริมวิทยาศาสตร์ในโรงเรียนมีความสัมพันธ์ทางบวกอย่างไม่มีนัยสำคัญทางสถิติกับ

ความสามารถด้านวิทยาศาสตร์ของนักเรียน และไม่ส่งผลให้มีการปรับปรุงการสอนวิทยาศาสตร์ของครู

7. นักเรียนที่มีความสามารถทางวิทยาศาสตร์สูงกว่าค่าเฉลี่ยนานาชาติแตกต่างจากนักเรียนที่มีความสามารถทางวิทยาศาสตร์ต่ำกว่าค่าเฉลี่ยนานาชาติในด้าน การเห็นคุณค่าของวิทยาศาสตร์ ความสุขในการเรียนวิทยาศาสตร์ เวลาที่ใช้ในการเรียนวิทยาศาสตร์ ความสนใจในวิทยาศาสตร์ และมโนทัศน์เกี่ยวกับวิทยาศาสตร์ แต่ความสัมพันธ์นี้มีการเปลี่ยนแปลงไปตามเพศของนักเรียน ที่ตั้งของโรงเรียน และขนาดโรงเรียน
8. ความมั่งคั่งของครอบครัวมีผลต่อความสนใจในวิทยาศาสตร์ ความสุขในการเรียนวิทยาศาสตร์ ความเชื่อมั่นในสมรรถภาพของตนเองในการเรียนวิทยาศาสตร์ มโนทัศน์เกี่ยวกับการวิทยาศาสตร์ และทรัพยากรทางการศึกษาที่บ้าน
9. ดัชนีทางเศรษฐกิจ สังคมและวัฒนธรรมของครอบครัวมีผลต่อจำนวนเวลาที่ใช้ศึกษาวิทยาศาสตร์ การเห็นคุณค่าของวิทยาศาสตร์ ความสุขในการเรียนวิทยาศาสตร์ ความเชื่อมั่นในสมรรถภาพของตนเองในการเรียนวิทยาศาสตร์ มโนทัศน์เกี่ยวกับการวิทยาศาสตร์ และทรัพยากรทางการศึกษาที่บ้าน
10. ปัจจัยที่เพิ่มความน่าจะเป็นในการได้คะแนนวิทยาศาสตร์สูงกว่าค่าเฉลี่ยนานาชาติได้แก่ โรงเรียนในเมือง โรงเรียนรัฐบาล การเห็นคุณค่าของวิทยาศาสตร์ ทรัพยากรทางการศึกษาที่บ้าน ความสนใจในวิทยาศาสตร์ ความสุขในการเรียนวิทยาศาสตร์ ความเชื่อมั่นในสมรรถภาพของตนเองในการเรียนวิทยาศาสตร์ ดัชนีทางเศรษฐกิจ สังคมและวัฒนธรรมของครอบครัว นักเรียนเพศชาย และจำนวนเวลาที่ใช้ศึกษาวิทยาศาสตร์ แต่ปัจจัยข้อจำกัดด้านทรัพยากรทางการศึกษา การสอนวิทยาศาสตร์ของครู มโนทัศน์เกี่ยวกับการวิทยาศาสตร์ ชั้นเรียนขนาดเล็ก และความมั่งคั่งของครอบครัวลดความน่าจะเป็นในการได้คะแนนวิทยาศาสตร์สูงกว่าค่าเฉลี่ยนานาชาติ

ผลงานวิจัยหลายเรื่อง ระบุว่าครูยังขาดความรู้ด้านการประเมินเพื่อพัฒนาการเรียนรู้ของผู้เรียน ทำให้ครูไม่สามารถพัฒนาการเรียนของนักเรียนเป็นรายบุคคลได้อย่างมีคุณภาพ นงลักษณ์ วิรัชชัย (2545) ศึกษาผลการดำเนินงานโครงการนำร่องด้านกระบวนการปฏิรูปเพื่อพัฒนาคุณภาพการเรียนรู้ พบว่า ครูยังจัดกระบวนการเรียนรู้ได้น้อย การประเมินการเรียนรู้แนวใหม่ที่ครูได้ทำระดับมาก ได้แก่ การใช้วิธีการหลากหลายในการประเมิน การแจ้งผลการประเมินให้ผู้ปกครองทราบ การบอกวัตถุประสงค์และวิธีการประเมินให้นักเรียนทราบตั้งแต่ต้นและการกำหนดเกณฑ์การประเมินร่วมกับนักเรียน ประเด็นที่ครูยอมรับว่าทำได้น้อย คือ การส่งเสริมให้นักเรียนประเมินการสอนของครู การวางระบบการประเมินทั้งโรงเรียน การวิเคราะห์จุดแข็งจุดอ่อนของผลการประเมินที่ใช้อยู่ การร่วมมือกับผู้ปกครองนำผลการประเมินไปใช้ประโยชน์ ผลการวิเคราะห์ที่น่าสังเกตคือ ประเด็นที่ครูทำพฤติกรรมน้อยได้แก่ การประเมินการเรียนรู้แนวใหม่ที่ให้นักเรียนและผู้ปกครองมีบทบาทในการ

ประเมินมากขึ้น นอกจากนี้ยังพบว่ามีประเด็นของระดับการแสดงพฤติกรรมของครูที่มีความแตกต่างกันมากด้วย

นอกจากนี้ ชินภัทร ภูมิรัตน (2544) ยังพบว่าวิทยาศาสตร์ศึกษาของไทยมีปัญหา 6 เรื่อง คือ ปัญหาหลักสูตร การจัดการเรียนการสอน ครู สื่อการเรียนรู้ การวัดและการประเมินผล และผลสัมฤทธิ์ทางการเรียนของผู้เรียน ดังสรุปได้ดังนี้

1. หลักสูตรมีปัญหาเรื่องความต่อเนื่องเชื่อมโยงระหว่างระดับการศึกษาและขาดความเชื่อมโยงกับชีวิตจริง

2. การจัดการเรียนการสอนยังเน้นการอธิบายและการสาธิตทำให้นักเรียนขาดกระบวนการคิดและการทดลองปฏิบัติจริง

3. ครูมีปัญหาในเชิงปริมาณและคุณภาพ และควรมีการเร่งพัฒนาครูประจำการให้มีคุณภาพด้านการสอน และมีปริมาณเพียงพอ

4. สื่อการสอนมีข้อจำกัด ขาดความหลากหลาย ขาดคุณภาพมาตรฐานและราคาที่เหมาะสม สื่อโทรทัศน์เพื่อการเรียนรู้และการฝึกอบรมครูยังไม่น่าสนใจเท่าที่ควร

5. การวัดและประเมินผลเน้นความรู้ ความจำ เน้นการเลือกคำตอบมากกว่าการวัดกระบวนการคิด ทำให้นักเรียนขาดทักษะการสื่อสาร การประเมินผลในสภาพจริงมีการดำเนินงานในวงจำกัด

6. ผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ของนักเรียนลดลงเรื่อยๆ ทุกวิชา ตั้งแต่วิชาเคมี ชีววิทยา ฟิสิกส์ และวิทยาศาสตร์กายภาพ

จากข้อจำกัดของครูด้านการใช้ผลการประเมินเพื่อปรับปรุงการเรียนของนักเรียน จึงมีความจำเป็นในการช่วยเหลือครูวิทยาศาสตร์ในรูปของการอบรมให้ความรู้ และฝึกให้ครูได้ปฏิบัติการออกแบบระบบและเครื่องมือการประเมินร่วมกันกับบุคลากรทางการศึกษาในระดับโรงเรียน และเขตพื้นที่การศึกษา การทำงานร่วมกันจะทำให้ผลการพัฒนาครูมีความยั่งยืน ครูและบุคลากรทางการศึกษาในระดับโรงเรียน และเขตพื้นที่การศึกษาสามารถทำการประเมินเพื่อพัฒนาการศึกษาร่วมกันในพื้นที่ของตนเองในอนาคตได้ อันจะเป็นส่วนหนึ่งของการช่วยให้การปฏิรูปการศึกษาประสบความสำเร็จได้

สรุป

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง นำมาสู่แนวคิดของการออกแบบการตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาตอนต้นด้วยระบบอินเทอร์เน็ตได้ว่า ระบบที่พัฒนาขึ้นควรมีคุณสมบัติในด้านความตรง ซึ่งเป็นคุณสมบัติในเชิงการวัด ประเด็นที่ควรตรวจสอบ คือ ความตรง ความเที่ยง ค่าเฉลี่ย ความแปรปรวนที่ควรมีค่าใกล้เคียงกันกับการทดสอบรูปแบบเดิมที่ใช้กันทั่วไป คือ การทดสอบด้วยการเขียนตอบ (paper-based testing) ในการตรวจสอบความเท่าเทียมกันของการทดสอบที่ต่างกัน นักวิชาการจึงดำเนินการเปรียบเทียบในประเด็นเหล่านี้ หนึ่งในการวิเคราะห์ที่สามารถบ่งบอกความเท่าเทียมกัน

ของการทดสอบที่ใช้รูปแบบต่างกัน คือ การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning หรือ DIF) ส่วนการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบข้อสอบจะทำให้ทราบความคลาดเคลื่อนมาตรฐานของการวัดของแบบทดสอบ และสารสนเทศของแบบทดสอบ (test information) ซึ่งเป็นดัชนีสะท้อนความเที่ยง เหมือนกับความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงพัฒนา โดยมีวัตถุประสงค์ เพื่อ 1) เพื่อสร้างข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น 2) เพื่อพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต 3) เพื่อตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต และ 4) เพื่อนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต ผู้วิจัยดำเนินการวิจัยดังต่อไปนี้

1. ประชากรและกลุ่มตัวอย่าง

ประชากรสำหรับการวิจัย คือ นักเรียนระดับมัธยมศึกษาปีที่ 3 ซึ่งเป็นกลุ่มเดียวกับนักเรียนที่เป็นเป้าหมายของการประเมิน PISA กลุ่มตัวอย่าง คือ นักเรียนระดับมัธยมศึกษาปีที่ 3 จำนวน 200 คน โดยสุ่มมาแบบเจาะจง โดยใช้เกณฑ์ คือ โรงเรียนต้องมีคอมพิวเตอร์ที่เชื่อมต่อกับอินเทอร์เน็ต เพื่อให้สามารถดำเนินการจัดสอบผ่านคอมพิวเตอร์ได้

2. เครื่องมือวิจัย

เครื่องมือวิจัยสำหรับการวิจัยเรื่องนี้ มี สองประเภทใหญ่ ๆ คือ ข้อสอบวิชาวิทยาศาสตร์ ซึ่งวัดความรู้และทักษะตามกรอบการประเมินนักเรียนระดับนานาชาติ (PISA) และเทคโนโลยีการทดสอบเพื่อใช้ในการจัดสอบที่พัฒนาขึ้น

3. การพัฒนาเครื่องมือวิจัย

การพัฒนาข้อสอบวัดการรู้วิทยาศาสตร์ และการพัฒนาระบบเทคโนโลยีการทดสอบเพื่อใช้ในการจัดสอบ วิธีดำเนินการวิจัยโดยละเอียดมีดังนี้

ขั้นที่ 1 การสร้างแบบวัดการรู้วิทยาศาสตร์ (scientific literacy)

ข้อสอบที่สร้างขึ้นเป็นข้อสอบเลือกตอบ (multiple-choice item) 17 ข้อ และข้อสอบเลือกตอบต่อเนื่อง (extended multiple-choice item) 3 ข้อ ซึ่งมีข้อสอบย่อย 2, 4, และ 4 ข้อ ตามลำดับ รวมเป็นข้อสอบจำนวน 27 ข้อ แต่ละข้อให้คะแนน 1 คะแนน ข้อสอบที่พัฒนาขึ้นแต่ละข้อวัดเนื้อหาความรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 โดยสมรรถนะที่ต้องการประเมินเป็นไปตามกรอบการประเมินการรู้วิทยาศาสตร์ของ PISA ดังตาราง 3.1 ดังนี้

1. ระบุประเด็นทางวิทยาศาสตร์ (identifying scientific issues) ประกอบด้วย การรู้ว่าประเด็นใดบ้างที่สามารถตรวจสอบได้อย่างวิทยาศาสตร์ การระบุคำสำคัญที่ต้องใช้ค้นหาข้อมูลทางวิทยาศาสตร์ และการรู้ลักษณะสำคัญของการตรวจสอบทางวิทยาศาสตร์ มีข้อสอบจำนวน 8 ข้อ

2. อธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์ (explaining phenomena scientifically) ประกอบด้วยการใช้ความรู้วิทยาศาสตร์ในสถานการณ์ที่กำหนด การอธิบายหรือแปลผลปรากฏการณ์อย่างวิทยาศาสตร์ และทำนายการเปลี่ยนแปลง การบอกถึงการบรรยาย อธิบาย และทำนายที่เหมาะสม มีข้อสอบจำนวน 12 ข้อ

3. ใช้หลักฐานทางวิทยาศาสตร์ (Using scientific evidence) ประกอบด้วย การแปลผลหลักฐานทางวิทยาศาสตร์ สรุป และสื่อสารให้เข้าใจได้ การระบุข้อตกลงเบื้องต้น หลักฐาน และเหตุผลเบื้องหลังข้อสรุป และการสะท้อนนัยของการนำผลการพัฒนาทางวิทยาศาสตร์และเทคโนโลยีไปใช้ในสังคม มีข้อสอบจำนวน 7 ข้อ

ตาราง 3.1 โครงสร้างการพัฒนาข้อสอบ

สมรรถนะ	จำนวนข้อ	ร้อยละ
ระบุประเด็นทางวิทยาศาสตร์	8	29.6
อธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์	12	44.4
ใช้หลักฐานทางวิทยาศาสตร์	7	25.9

ข้อสอบที่สร้างขึ้นจะมีลักษณะต่างจากการประเมินด้วยแบบทดสอบแบบกระดาษ กล่าวคือผู้สอบสามารถคลิ๊กตอบผ่านคอมพิวเตอร์ได้

ขั้นที่ 2 ตรวจสอบคุณภาพข้อสอบและแบบทดสอบ

การตรวจสอบคุณภาพหลังสร้างข้อสอบเสร็จแล้ว ใช้การตรวจสอบคุณภาพข้อสอบโดยผู้ทรงคุณวุฒิด้านเนื้อหาจำนวน 5 คน เพื่อตรวจสอบความตรงเชิงเนื้อหา โดยผลการวิเคราะห์ค่าความสอดคล้องกับเนื้อหา พบว่าข้อสอบมีค่า IOC ตั้งแต่ 0.6-1.0 จากนั้นนำไปทดลองใช้กับนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 3 จำนวน 30 คน เพื่อตรวจสอบความเที่ยงของแบบทดสอบ โดยวิเคราะห์ค่าความเที่ยง KR-21 พบว่ามีความเที่ยง 0.7 ผู้วิจัยนำผลการตรวจสอบมาปรับปรุงข้อสอบที่มีอำนาจจำแนกต่ำให้มีคุณภาพต่อไป

4. การกำหนดคะแนนจุดตัด

การวิจัยครั้งนี้ใช้วิธีบุ๊คมาร์ค (Bookmark Method) ในการกำหนดจุดตัด โดยมีขั้นตอนดังนี้

4.1 นำแบบทดสอบไปสอบนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 6 ห้องเรียน โดยมีจำนวนนักเรียนทั้งหมด 200 คน จากนั้นนำผลการทดสอบมาวิเคราะห์ข้อสอบด้วยโมเดลการตอบข้อสอบ (item response theory) เพื่อต้องการได้ค่าความยาก (difficulty) อำนาจจำแนก (a) และค่าการเดา (guessing) เพื่อนำไปกำหนดจุดตัด (passing score) ด้วยวิธีบุ๊คมาร์ค (Bookmark Method) สำหรับใช้แปลผลความสามารถจากการทดสอบผ่านอินเทอร์เน็ต โดยมีผู้เชี่ยวชาญเข้าร่วมในการกำหนด

จุดตัด 5 คน ตารางที่ 3.2 แสดงว่าความยาก อำนาจจำแนก และการเดาของข้อสอบ 27 ข้อ โดยค่า a มีค่า 1.15 ถึง 2.16 ค่า b มีค่า 0.40-1.37 ส่วนค่า c มีค่า 0.27-0.36

ตารางที่ 3.2 ความยาก อำนาจจำแนก และการเดาของข้อสอบ 27

	a	b	c
ข้อที่ 1	1.49	0.69	0.29
ข้อที่ 2	1.73	0.69	0.30
ข้อที่ 3	1.35	0.73	0.29
ข้อที่ 4	1.53	0.93	0.32
ข้อที่ 5	1.72	0.89	0.34
ข้อที่ 6	1.67	0.78	0.27
ข้อที่ 7	1.88	0.70	0.28
ข้อที่ 8	1.57	0.73	0.31
ข้อที่ 9	1.45	0.54	0.33
ข้อที่ 10	**1.47**	**0.83**	**0.35**
ข้อที่ 11	1.15	1.37	0.32
ข้อที่ 12	1.70	0.89	0.32
ข้อที่ 13	2.13	0.69	0.30
ข้อที่ 14	1.62	0.77	0.30
ข้อที่ 15	1.71	0.75	0.28
ข้อที่ 16	1.42	0.91	0.29
ข้อที่ 17	1.47	0.92	0.29
ข้อที่ 18	2.16	0.40	0.31
ข้อที่ 19	1.77	0.69	0.27
ข้อที่ 20	1.52	0.91	0.30
ข้อที่ 21	2.14	0.55	0.28
ข้อที่ 22	1.59	0.78	0.25
ข้อที่ 23	1.55	0.98	0.36
ข้อที่ 24	1.83	0.78	0.31
ข้อที่ 25	1.43	1.08	0.27
ข้อที่ 26	1.70	0.89	0.30
ข้อที่ 27	1.80	0.50	0.26

4.2 ผู้วิจัยนำข้อสอบมาจัดฉบับใหม่ โดยเรียงข้อใหม่ตามค่า b จากน้อยไปมาก และให้ผู้เชี่ยวชาญ 5 ท่าน พิจารณาข้อสอบทีละข้อ และช่วยระบุว่าข้อใดเป็นข้อที่ผู้สอบมีโอกาสตอบถูกประมาณร้อยละ 67 ผลการดำเนินการพบว่า ข้อ 10 เป็นข้อที่ผู้เชี่ยวชาญเห็นว่าเป็นข้อที่ผู้สอบมี

โอกาสตอบถูกประมาณ ร้อยละ 67 ดังนั้นค่าความยาก อำนาจจำแนก และการเดาของข้อนี้จะนำไปคำนวณจุดตัด (θ) ของการรายงานผลการสอบ ด้วยการหาจุดตัดด้วยวิธีบุ๊คมาร์ค (Bookmark Method)

4.3 คำนวณค่าคะแนนจุดตัด หรือ θ ของสมการนี้

$$0.67 = c + \frac{\exp[a(\theta - b)]}{1 + \exp[a(\theta - b)]}$$

ผลการคำนวณค่าคะแนนจุดตัด θ ได้เท่ากับ .34 ซึ่งมีค่าเท่ากับ 534 คะแนนในช่วงสเกลของการประเมินนักเรียนระดับนานาชาติ (PISA)

5. การพัฒนาระบบการทดสอบ

การพัฒนาระบบการทดสอบมีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 พัฒนาระบบเทคโนโลยีการทดสอบเพื่อใช้ในการจัดสอบ ด้วยโปรแกรม PHP ซึ่งประกอบด้วยระบบการนำข้อสอบเข้าฐานข้อมูล การสร้าง/ปรับข้อสอบเพื่อเพิ่มเติมภาพ แอนิเมชั่น เสียง และภาพเคลื่อนไหว ระบบการจัดสอบผ่านอินเทอร์เน็ต และรายงานผลการทดสอบ ซึ่งผู้ทรงคุณวุฒิ 5 ท่าน ตรวจสอบคุณภาพของระบบและเสนอแนะให้มีการปรับการนำเสนอข้อสอบในระหว่างการทดสอบผ่านคอมพิวเตอร์ให้ชัดเจนขึ้น

ขั้นที่ 2 นำข้อสอบที่พัฒนาขึ้น เข้าระบบฐานข้อมูล โดยการนำเข้าข้อสอบรายข้อ และดำเนินการเติมภาพ เสียง ภาพเคลื่อนไหว และแอนิเมชั่นเพิ่มเติมในข้อสอบที่ต้องการ

ขั้นที่ 3 นำระบบไปทดสอบผ่านอินเทอร์เน็ตกับนักเรียนจำนวน 200 คน เพื่อตรวจสอบความเหมาะสมของการทดสอบในด้านความตรง และความเที่ยง ซึ่งวิเคราะห์ความเที่ยงได้ 0.77 และวิเคราะห์ความตรงตามเกณฑ์สัมพันธ์ด้วยการหาความสัมพันธ์ระหว่างคะแนนสอบ (θ) กับเจตคติกับวิทยาศาสตร์ได้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันเท่ากับ .21 และมีนัยสำคัญทางสถิติ (p=.029)

ขั้นที่ 4 นำแบบทดสอบแบบที่ทำผ่านกระดาษ (paper & pencil) ไปทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 100 เพื่อนำผลการทดสอบมาวิเคราะห์เปรียบเทียบคุณภาพกับแบบทดสอบด้วยคอมพิวเตอร์ โดยเปรียบเทียบสารสนเทศของแบบทดสอบ ความคลาดเคลื่อนมาตรฐานของการวัด ความเที่ยง

4. การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลประกอบด้วยการวิเคราะห์เพื่อเปรียบเทียบแบบทดสอบที่ทำผ่านคอมพิวเตอร์ และแบบทดสอบแบบที่ทำผ่านกระดาษ (paper & pencil) โดยเปรียบเทียบด้านต่อไปนี้

1. ความเที่ยง โดยใช้การวิเคราะห์ดัชนีความเที่ยงด้วยวิธี KR-21

2) การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning) ด้วยการวิเคราะห์การเปรียบเทียบค่าความยากขอแบบทดสอบทั้งสองประเภทด้วยการวิเคราะห์การตอบข้อสอบ (item response theory)

3. การเปรียบเทียบคะแนนสอบระหว่างกลุ่มผู้สอบ IBT และ PBT ด้วยสถิติ t-test

4. เปรียบเทียบความเที่ยงแบบทดสอบด้วยทฤษฎีการตอบข้อสอบ โดยพิจารณาเปรียบเทียบสารสนเทศของแบบทดสอบ และความคลอดเคลื่อนมาตรฐานของการวัด

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยนี้มุ่งพัฒนาระบบการประเมินการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ต เพื่อส่งเสริมการเรียนรู้ขั้นสูงในวิชาวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น โดยมีวัตถุประสงค์ย่อยดังนี้

1. เพื่อสร้างข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น
2. เพื่อพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต
3. เพื่อตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต
4. เพื่อนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต

ในการนำเสนอผลการวิจัย ผู้วิจัยนำเสนอผลการวิจัยตามวัตถุประสงค์ของการวิจัย โดยแบ่งออกเป็น 4 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์คุณภาพข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น

ตอนที่ 2 ผลการพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต

ตอนที่ 3 ผลการตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต

ตอนที่ 4 ผลการนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต

ซึ่งในการนำเสนอ ผู้วิจัยใช้คำว่า IBT แทนการทดสอบผ่านคอมพิวเตอร์ (internet-based testing) และ PBT แทนคำว่า การทดสอบแบบเขียนตอบด้วยกระดาษ (paper-based testing)

ตอนที่ 1 ผลการวิเคราะห์คุณภาพข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น

ผลการวิเคราะห์คุณภาพด้านความเที่ยงของแบบทดสอบ IBT และ PBT ด้านความเที่ยง (KR-21) พบว่าแบบทดสอบ IBT เท่ากับ .76 และความเที่ยงของแบบ PBT เท่ากับ .68 ส่วนความยากและอำนาจจำแนกของแบบทดสอบทั้งสองฉบับ อยู่ในตารางที่ 4.1 โดยความยากของแบบทดสอบ IBT มีค่าระหว่าง .20 ถึง .78 (M=.44, SD=.16) ค่าความยากของแบบทดสอบ PBT มีค่าระหว่าง .15 ถึง .82 (M=.43, SD=.18) ค่าอำนาจจำแนกของแบบทดสอบ IBT มีค่า -.23 ถึง .42 (M=.12, SD=.18) ค่าอำนาจจำแนกของแบบทดสอบ PBT มีค่า -.16 ถึง .32 (M=.09, SD=.11)

ตารางที่ 4.1 ความยากและอำนาจจำแนกของแบบทดสอบ IBT และ PBT

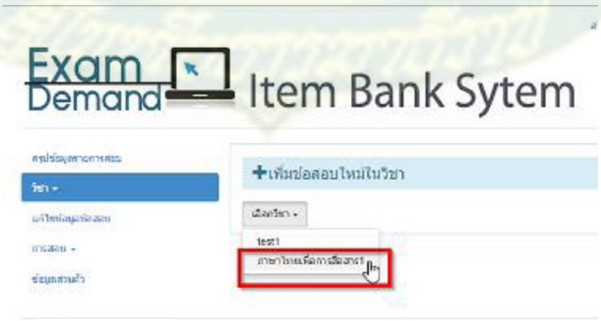
ข้อที่	ความยาก		อำนาจจำแนก	
	IBT	PBT	IBT	PBT
ข้อ1	.34	.28	-.06	.12
ข้อ2	.36	.41	-.06	.07
ข้อ3	.25	.30	.21	.12
ข้อ4	.21	.15	-.23	.04
ข้อ5	.78	.80	-.01	.07
ข้อ6	.78	.82	.28	.20
ข้อ7	.58	.61	.36	.32
ข้อ8	.36	.37	.42	.05
ข้อ9	.39	.36	.22	.05
ข้อ10	.20	.20	-.23	-.07
ข้อ11	.52	.68	.30	.06
ข้อ12	.54	.59	.11	.18
ข้อ13	.55	.33	.17	.01
ข้อ14	.61	.51	.24	.16
ข้อ15	.28	.34	.03	.06
ข้อ16	.64	.63	.23	.18
ข้อ17	.60	.65	.32	.17

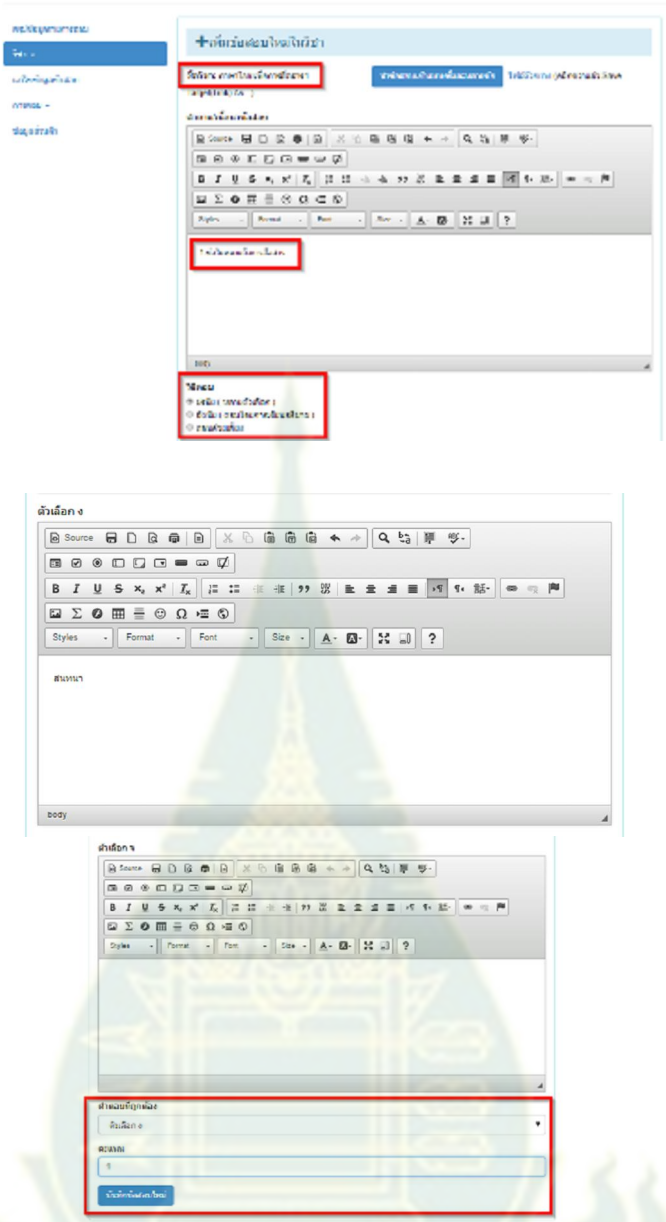
ข้อที่	ความยาก		อำนาจจำแนก	
	IBT	PBT	IBT	PBT
ข้อ18	.51	.39	-.05	-.09
ข้อ19	.46	.44	.12	.04
ข้อ20	.26	.26	.06	.03
ข้อ21	.42	.49	-.22	.09
ข้อ22	.29	.30	.26	-.02
ข้อ23	.51	.33	.11	.19
ข้อ24	.36	.47	.17	.27
ข้อ25	.21	.18	-.11	-.16
ข้อ26	.42	.36	.17	.17
ข้อ27	.36	.29	.32	.10
เฉลี่ย	.44	.43	.12	.09
SD	0.16	0.18	0.18	0.11

ตอนที่ 2 ผลการพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต

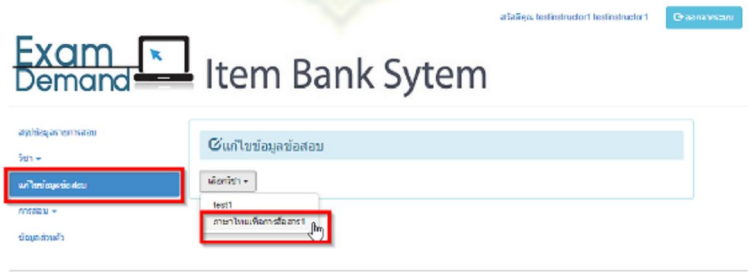
ระบบที่พัฒนาขึ้นเป็นระบบการสอบคอมพิวเตอร์โดยผ่านอินเทอร์เน็ต ผู้วิจัยพัฒนาให้มีองค์ประกอบย่อยดังนี้

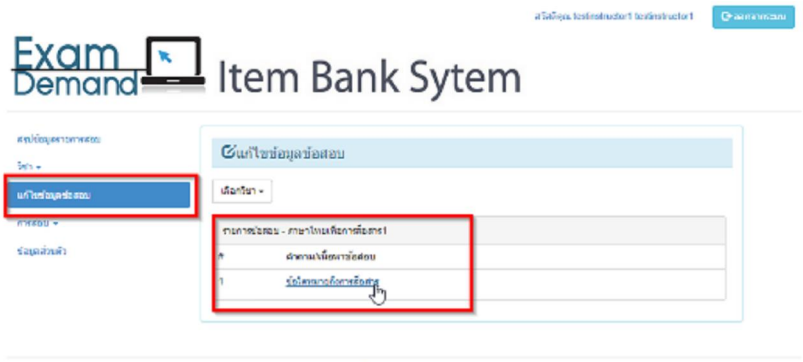
1. ระบบการนำเข้าข้อสอบสู่ระบบ ระบบนี้เป็นระบบการนำเข้าข้อสอบ โดยต้องมีการพิมพ์ข้อสอบเข้าสู่ระบบรายข้อ ข้อสอบที่นำเข้าจะนำไปจัดเป็นแบบทดสอบเพื่อใช้เป็นแบบทดสอบสำหรับการทดสอบ ข้อสอบที่นำเข้าเป็นได้ทั้งข้อความ ภาพ และเสียง เพื่อให้สามารถทำการวัดความสามารถทางวิทยาศาสตร์ของผู้สอบได้ครอบคลุม





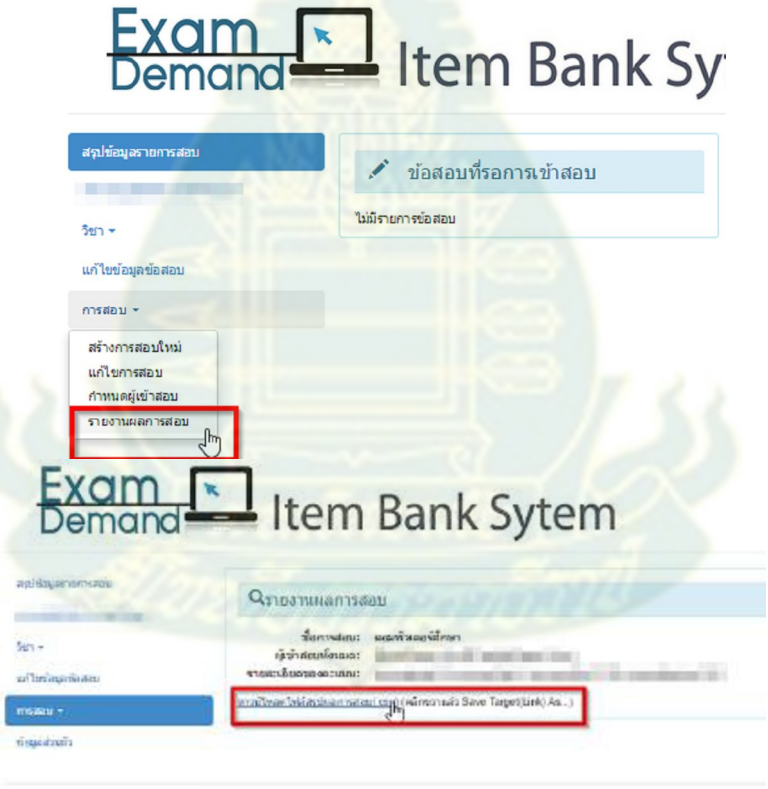
2. ระบบระบบการเขียน/ปรับข้อสอบ ระบบนี้เป็นระบบที่พัฒนาขึ้นมาเพื่อให้ผู้ใช้ระบบสามารถปรับปรุงแก้ไขข้อสอบให้ถูกต้อง และตรงตามความต้องการของผู้ใช้





3. ระบบระบบการสอบผ่านอินเทอร์เน็ต ระบบการสอบเป็นระบบที่ประกอบด้วยข้อสอบและสื่อการนำเสนอ คือ การนำเสนอข้อสอบผ่านหน้าจอคอมพิวเตอร์ ผู้สอบจะต้องมีรหัสเข้าสอบเพื่อใช้จำแนกผลการสอบ และรายงานผลการสอบเป็นรายบุคคล ซึ่งผู้สอบสามารถเข้าสู่ระบบได้ทั้งคอมพิวเตอร์ และโทรศัพท์มือถือ

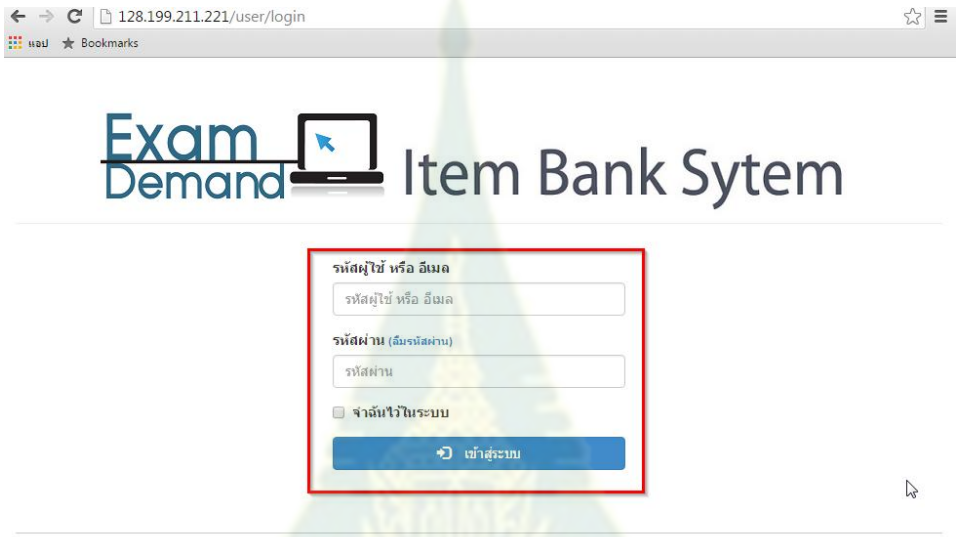
4. ระบบการรายงานผลการสอบ ระบบรายงานผลเป็นระบบที่วิเคราะห์ผลการทดสอบรายบุคคล โดยรายงานคะแนนสอบในรูปของคะแนนดิบ และผลการสอบผ่าน หรือไม่ผ่าน



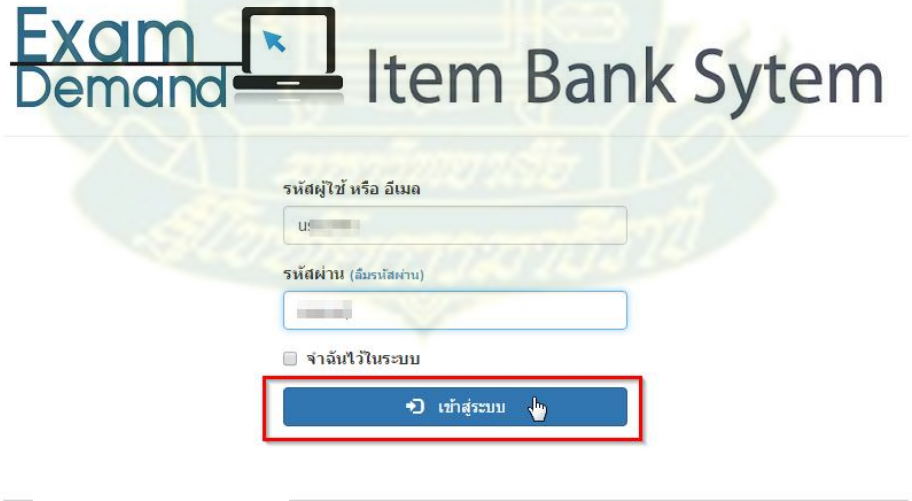
สำหรับขั้นตอนการเข้าทดสอบผ่านระบบอินเทอร์เน็ต มีดังนี้

1. นักเรียนเปิดเว็บไซต์ด้วย Web Browser -- > Google Chrome (ทำงานได้ดี) หรือ Mozilla Firefox

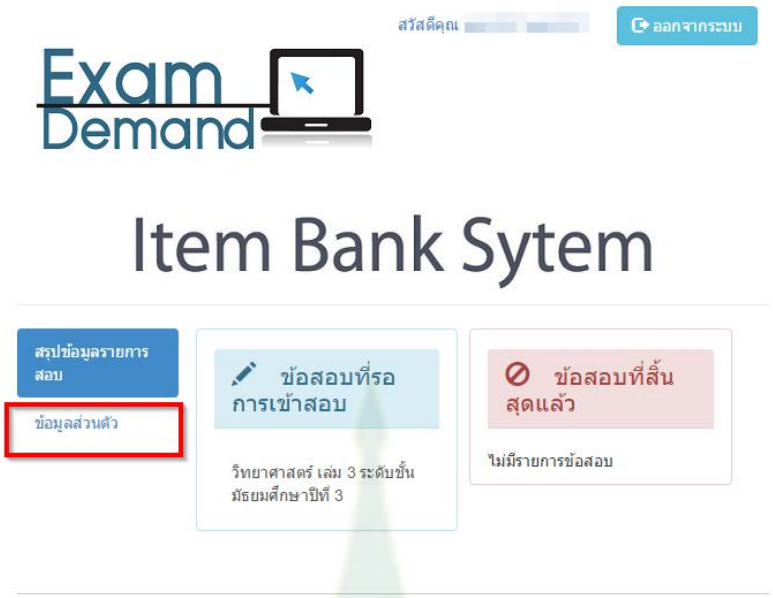
2. ให้นักเรียนเข้าที่ URL : http://128.199.211.221 หรือ 128.199.211.221 ก็จะปรากฏหน้าต่างตามภาพ



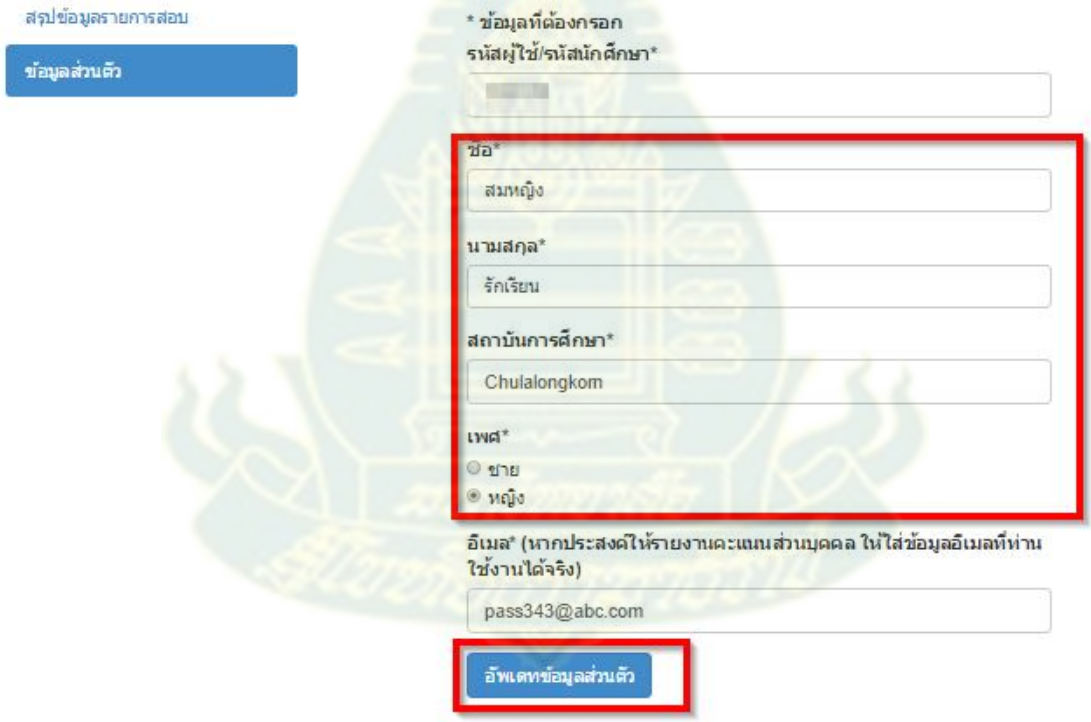
3. กรอกข้อมูล รหัสผู้ใช้ และ รหัสผ่าน ตามภาพ แล้วกดปุ่มเข้าสู่ระบบ



4. นักเรียนไปแก้ข้อมูลส่วนตัว ตามภาพ

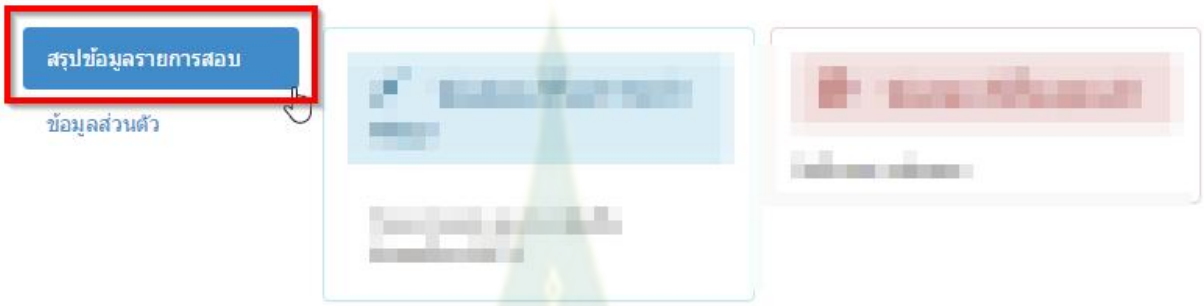


5. เมื่อเข้าสู่หน้าต่างนี้แล้ว ให้นักเรียนแก้ไข **ชื่อ นามสกุล สถาบันการศึกษา และ เพศ** เสร็จแล้ว กดปุ่ม “อัพเดตข้อมูลส่วนตัว” ตามภาพ



6. เมื่อกดปุ่ม “อัพเดตข้อมูลส่วนตัว” เสร็จแล้ว ที่หน้าต่างเดิมจะขึ้นแถบสีเขียว ตามภาพ

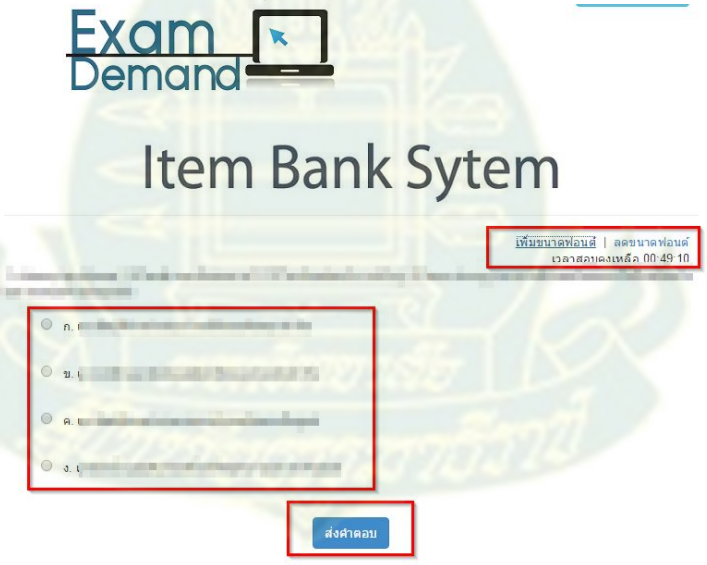
7. หลังจากนั้นให้คลิกที่ เมนู “สรุปข้อมูลรายการสอบ” ทางด้านซ้ายมือ ตามภาพ



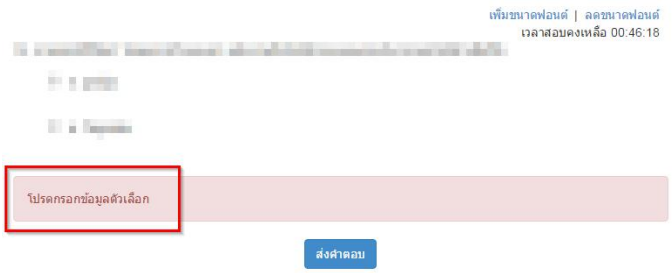
8. นักเรียนคลิกที่ ชื่อวิชา วิทยาศาสตร์ เล่ม 3 ระดับชั้นมัธยมศึกษาปีที่ 3 เพื่อเข้าทดสอบ

9. นักเรียนอ่านคำแนะนำ รายละเอียดการสอบต่างๆ ก่อนการเข้าใช้ แล้วคลิกปุ่ม “เริ่มทำข้อสอบ”

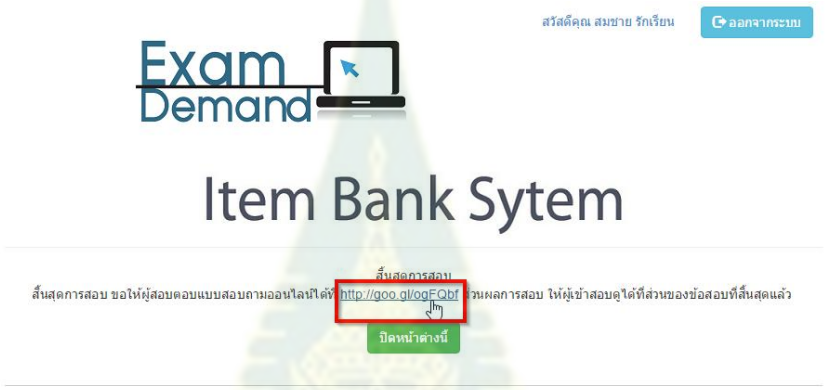
10. เมื่อคลิกปุ่มเริ่มทำข้อสอบ นักเรียนจะได้ทำข้อสอบไปเรื่อยๆ ตามจำนวนข้อ นักเรียนเลือกตอบตัวเลือกที่ต้องการ เมื่อเลือกเสร็จแล้วให้กดปุ่ม “ส่งคำตอบ”



11. หากข้อใด นักเรียนไม่ได้กดปุ่มเลือกตัวเลือกใดๆเลย เมื่อกดปุ่มส่งคำตอบ จะปรากฏข้อความให้เราเลือกข้อมูลตัวเลือก ตามภาพ

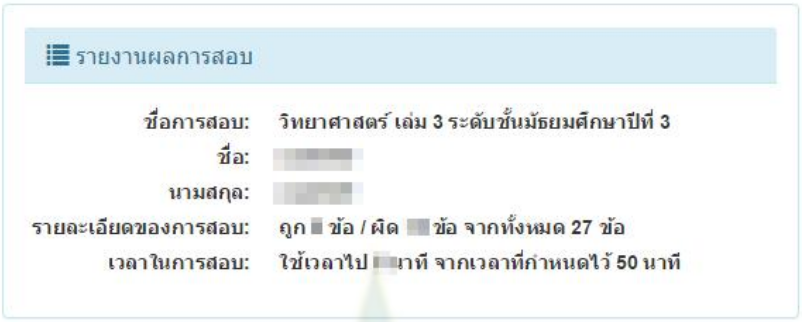


12. เมื่อทำข้อสอบครบทุกข้อแล้ว **<u>ให้คลิกเข้าไปทำแบบสอบถามออนไลน์ ตามลิงค์ที่ให้ไว้</u>** ตามภาพ



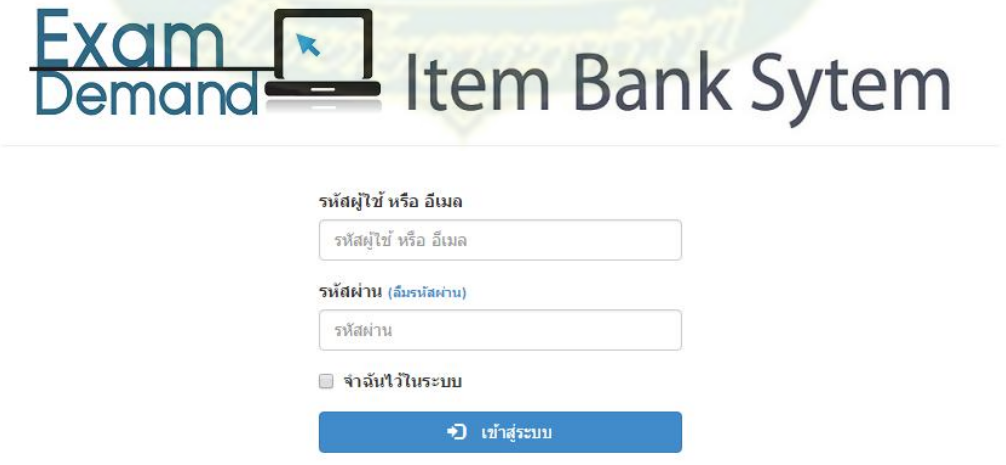
13. ให้นักเรียนตอบแบบสอบถามออนไลน์ **<u>ทุกข้อ แล้วกดปุ่ม “ส่ง”</u>** ตามภาพ
14. หลังจากนั้นกลับมาที่หน้าเมนู ให้กดปุ่ม Reflesh หรือ F5 หน้าจอของ Web Browser นักเรียนก็จะสามารถเข้าไปดูคะแนนวิชาที่ตนเองสอบได้ ตามภาพ หลังจากนั้นนักเรียนสามารถเลือกในส่วนของ “ข้อสอบที่สิ้นสุดแล้ว” เลือกวิชาที่ต้องการ เพื่อดูผลการสอบ

15. ระบบจะทำการคำนวณคะแนนผลการสอบออกมา โดยสรุปตามภาพ



16. หลังจากนั้นนักเรียนคลิกปุ่ม “ออกจากระบบ” ตามภาพ

17. เมื่อออกมาพบหน้าต่างนี้ แสดงว่านักเรียนได้ออกจากระบบเป็นที่เรียบร้อยแล้ว



ตอนที่ 3 ผลการตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต

การตรวจสอบคุณภาพของการทดสอบผ่านอินเทอร์เน็ตในครั้งนี้ มุ่งวิเคราะห์ความเท่าเทียมกัน (comparability) ในเชิงคุณสมบัติทางการวัดทางจิต (psychometric property) ของระบบการทดสอบด้วยอินเทอร์เน็ต (internet-based testing, IBT) กับระบบการทดสอบด้วยการเขียนตอบ (paper-based testing, PBT)

ตอนที่ 3.1 ภูมิหลังของผู้สอบ

ตารางที่ 4.2 แสดงภูมิหลังของกลุ่มผู้สอบทั้งสองแบบ โดยมีผู้เข้าสอบระบบ IBT จำนวน 148 คน และผู้เข้าสอบกลุ่ม IBT จำนวน 157 คน กลุ่ม IBT มีจำนวนนักเรียนชายมากกกว่าหญิง แต่กลุ่ม PBT มีจำนวนนักเรียนหญิงมากกกว่าชาย

ตาราง 4.2 ภูมิหลังของผู้ทำแบบทดสอบ

	เพศ	ความถี่	ร้อยละ
IBT	ชาย	68	54.1
	หญิง	80	45.9
	รวม	148	100
PBT	ชาย	57	36.3
	หญิง	100	63.7
	รวม	157	100

ตารางที่ 4.3 นำเสนอคะแนนผลการทดสอบในรูปคะแนนดิบของผู้เข้าสอบกลุ่ม IBT และ PBT ซึ่ง กลุ่ม IBT มีค่าเฉลี่ย 11.7 คะแนน SD=3.35 ส่วนกลุ่ม PBT มีค่าเฉลี่ย 11.52 SD=3.02

ตาราง 4.3 สถิติพื้นฐานของคะแนนสอบ (คะแนนดิบ)

ประเภท	Min	Max	M	SD
IBT	5.00	21.00	11.70	3.35
PBT	3.00	21.00	11.52	3.02

ตอนที่ 3.2 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning, DIF)

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นการเปรียบเทียบความยากของของข้อสอบจากแบบทดสอบ IBT และ PBT ด้วยสถิติที (t-test) นำเสนอในตารางที่ 4.4 ผลการวิเคราะห์พบว่ามีข้อสอบจำนวน 23 ข้อ จาก 24 ข้อ ที่ค่า t ไม่มีนัยสำคัญทางสถิติ มีข้อสอบ 1 ข้อ คือ ข้อที่ 13 มีค่า t=-2.18 และมีนัยสำคัญทางสถิติ ซึ่งเป็นข้อที่มีความยากสูงสำหรับการทดสอบแบบ PBT แต่ไม่ค่อยยากสำหรับผู้สอบที่สอบแบบ IBT

ตาราง 4.4 ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่ม IBT และ PBT

ข้อที่	อำนาจจำแนก (a)		ความยาก (b)		การเดา (c)		t
	IBT	PBT	IBT	PBT	IBT	PBT	
ITEM0001	0.21	0.21	3.35	4.44	0.04	0.04	-0.50
se	0.07	0.07	1.38	1.67	0.03	0.03	
ITEM0002	0.27	0.27	2.65	1.93	0.06	0.06	0.49
	0.09	0.09	1.14	0.95	0.05	0.05	
ITEM0003	0.53	0.53	2.44	1.95	0.05	0.05	0.50
se	0.16	0.16	0.71	0.66	0.04	0.04	
ITEM0005	0.25	0.25	-3.10	-3.53	0.09	0.09	0.22
se	0.08	0.08	1.35	1.46	0.07	0.07	
ITEM0006	1.19	1.19	-1.36	-1.42	0.06	0.06	0.15
se	0.33	0.33	0.27	0.34	0.05	0.05	
ITEM0007	1.14	1.14	-0.33	-0.36	0.05	0.05	0.11
se	0.24	0.24	0.22	0.21	0.04	0.04	
ITEM0008	0.93	0.93	0.99	0.82	0.04	0.04	0.43
se	0.21	0.21	0.30	0.25	0.03	0.03	
ITEM0009	0.57	0.57	1.30	1.50	0.05	0.05	-0.29
se	0.16	0.16	0.47	0.53	0.04	0.04	
ITEM0011	0.52	0.52	0.17	-1.22	0.09	0.09	1.84
se	0.14	0.14	0.51	0.56	0.07	0.07	
ITEM0012	0.39	0.39	0.05	-0.54	0.08	0.08	0.71
se	0.11	0.11	0.59	0.58	0.07	0.07	
ITEM0013	0.25	0.25	-0.10	3.46	0.07	0.07	-2.18*
se	0.08	0.08	0.78	1.43	0.05	0.05	
ITEM0014	0.47	0.47	-0.38	0.27	0.08	0.08	-0.91
se	0.14	0.14	0.51	0.49	0.07	0.07	
ITEM0015	0.26	0.26	3.92	2.97	0.04	0.04	0.50
se	0.08	0.08	1.47	1.21	0.04	0.04	
ITEM0016	0.79	0.79	-0.66	-0.61	0.07	0.07	-0.12

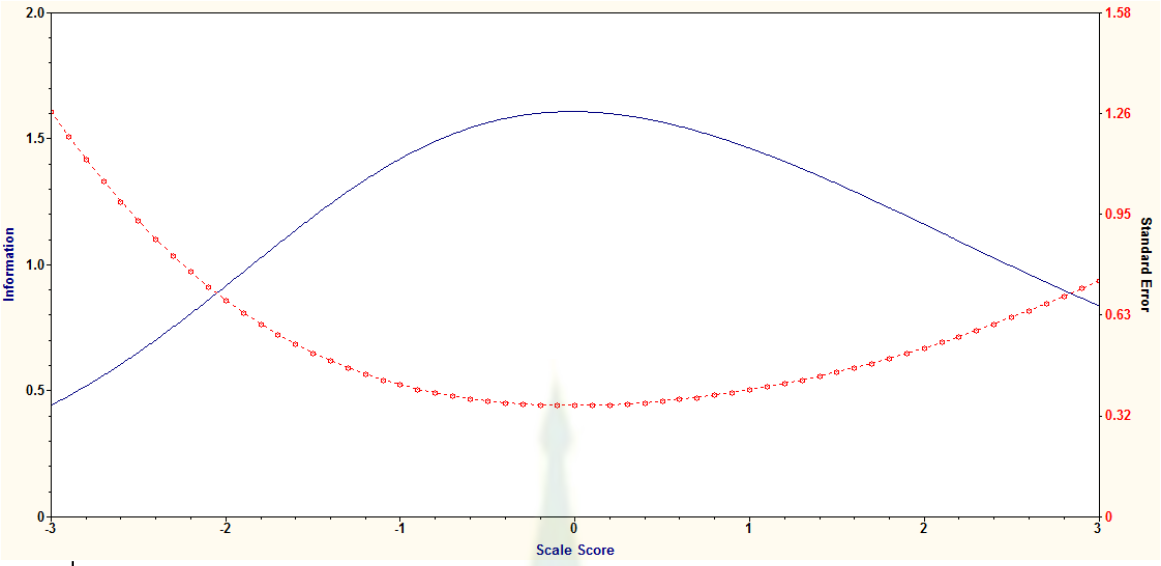
ข้อที่	อำนาจจำแนก (a)		ความยาก (b)		การเดา (c)		t
	IBT	PBT	IBT	PBT	IBT	PBT	
se	0.18	0.18	0.32	0.31	0.06	0.06	
ITEM0017	0.91	0.91	-0.46	-0.61	0.07	0.07	0.39
se	0.20	0.20	0.28	0.28	0.05	0.05	
ITEM0018	0.18	0.18	0.94	3.17	0.06	0.06	-1.23
se	0.06	0.06	1.05	1.48	0.05	0.05	
ITEM0019	0.33	0.33	1.24	1.25	0.08	0.08	-0.01
se	0.10	0.10	0.77	0.76	0.06	0.06	
ITEM0020	0.33	0.33	3.76	3.67	0.04	0.04	0.05
se	0.11	0.11	1.30	1.33	0.03	0.03	
ITEM0021	0.17	0.17	2.51	1.17	0.07	0.07	0.78
se	0.06	0.06	1.34	1.09	0.05	0.05	
ITEM0022	0.50	0.50	2.28	2.10	0.06	0.06	0.18
se	0.15	0.15	0.70	0.70	0.04	0.04	
ITEM0023	0.43	0.43	0.36	2.17	0.07	0.07	-1.88
se	0.12	0.12	0.51	0.82	0.05	0.05	
ITEM0024	0.69	0.69	1.26	0.39	0.06	0.06	1.63
se	0.19	0.19	0.43	0.32	0.05	0.05	
ITEM0026	0.60	0.60	0.97	1.38	0.07	0.07	-0.60
se	0.17	0.17	0.46	0.51	0.06	0.06	
ITEM0027	0.84	0.84	1.17	1.57	0.08	0.08	-0.69
se	0.26	0.26	0.37	0.45	0.06	0.06	

หมายเหตุ ข้อที่ 4, 10, และ 25 มีค่าสถิติไม่ผ่านเกณฑ์ โปรแกรม Bilog ไม่รายงานผลการวิเคราะห์

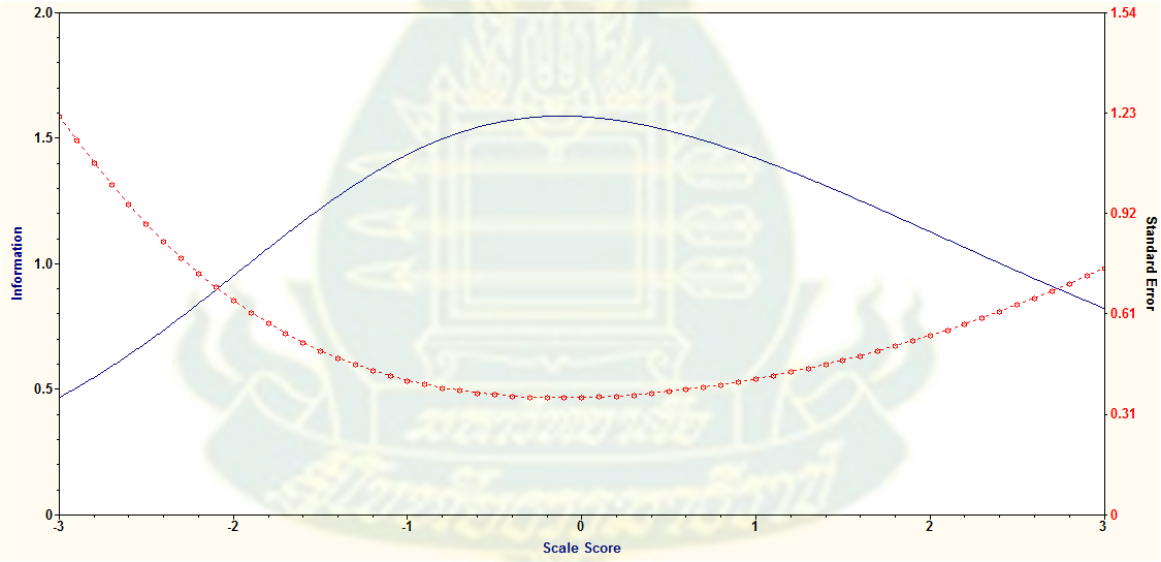
* p <.05

ตอนที่ 3.3 ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการวัด (standard error) สารสนเทศของแบบทดสอบ (information) ของแบบทสอบแบบ IBT และ PBT

ผลการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการวัด (standard error) สารสนเทศของแบบทดสอบ (information) ของแบบทสอบแบบ IBT และ PBT อยู่ในภาพที่ 4.1 และ 4.2 ซึ่งจะเห็นว่าความคลาดเคลื่อนมาตรฐานของการวัด และสารสนเทศของแบบทดสอบทั้งสองแบบมีค่า และรูปแบบลักษณะใกล้เคียงกัน



ภาพที่ 4.1 สารสนเทศของแบบทดสอบ IBT



ภาพที่ 4.2 สารสนเทศของแบบทดสอบ PBT

ตอนที่ 3.4 ผลการเปรียบเทียบคะแนนสอบของผู้สอบกลุ่ม IBT กับกลุ่ม PBT

ตารางที่ 4.5 แสดงผลการเปรียบเทียบคะแนนสอบของผู้สอบกลุ่ม IBT กับกลุ่ม PBT ด้วยสถิติ t-t-test จะเห็นว่ากลุ่ม IBT มีค่าเฉลี่ย 500.06, SD=100.09 กลุ่ม PBT มีค่าเฉลี่ย 499.94, SD=91.00 ซึ่งไม่ต่างกันอย่างมีนัยสำคัญทางสถิติ (t(305), =1.08, p=.28)

ตารางที่ 4.5 ผลการเปรียบเทียบคะแนนสอบของแบบทดสอบ IBT และ PBT

กลุ่ม	n	M	SD	t	df	p	Mean Difference	95% Confidence	
								Lower	Upper
IBT	148	500.06	100.09	1.08	305	.28	.12	-.10	.35
PBT	159	499.94	91.00						

Levene test=1.314, p=.253

ตอนที่ 4 ผลการนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต

ผู้วิจัยได้นำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปพัฒนาไว้ในรูปแบบการทดสอบผ่านอินเทอร์เน็ตเพื่อให้ครูที่ประสงค์ให้นักเรียนเข้าใช้ระบบการทดสอบ ซึ่งมีแจ้งความประสงค์ให้นักเรียนใช้ 46 คน โดยมีนักเรียนใช้ระบบจำนวน 675 คน นักเรียนที่ใช้ระบบมีความสนใจในการทดสอบ และผลการสัมภาษณ์ครูจำนวน 2 คน พบว่า ระบบที่พัฒนาขึ้นมีประโยชน์ต่อครูในด้านการประเมินระหว่างเรียน ครูสามารถใช้การประเมินนี้เพื่อประเมินการเรียนรู้ของนักเรียน และการทดสอบผ่านคอมพิวเตอร์ช่วยให้นักเรียนมีความกระตือรือร้นในการทดสอบมากกว่าการสอบแบบทั่วไป นอกนี้การทดสอบผ่านคอมพิวเตอร์ยังช่วยพัฒนาทักษะการใช้คอมพิวเตอร์ของนักเรียนด้วย

บทที่ 5
สรุปผล และอภิปรายผลการวิจัย

การวิจัยนี้มุ่งพัฒนาระบบการประเมินการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ต เพื่อส่งเสริมการเรียนรู้ขั้นสูงในวิชาวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น โดยมีวัตถุประสงค์ย่อยดังนี้

1. เพื่อสร้างข้อสอบเพื่อวัดการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น
2. เพื่อพัฒนาระบบประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต
3. เพื่อตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต
4. เพื่อนำระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต

สรุปผลการวิจัย

การดำเนินการวิจัยตามวัตถุประสงค์ทั้ง 4 ข้อ ได้ผลการวิจัยสรุปได้ดังนี้

1. แบบทดสอบการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นเป็นข้อสอบเลือกตอบ (multiple-choice item) 17 ข้อ และข้อสอบเลือกตอบต่อเนื่อง (extended multiple-choice item) 3 ข้อ ซึ่งมีข้อสอบย่อย 2, 4, และ 4 ข้อตามลำดับ รวมเป็นข้อสอบจำนวน 27 ข้อ แต่ละข้อให้คะแนน 1 คะแนน ข้อสอบที่พัฒนาขึ้นแต่ละข้อวัดเนื้อหาความรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 โดยสมรรถนะที่ต้องการประเมินเป็นไปตามกรอบการประเมินการรู้วิทยาศาสตร์ของ PISA คือ ระบุประเด็นทางวิทยาศาสตร์ (identifying scientific issues) การอธิบายปรากฏการณ์ต่างๆ อย่างวิทยาศาสตร์ (explaining phenomena scientifically) และการใช้หลักฐานทางวิทยาศาสตร์ (Using scientific evidence) ข้อสอบความตรงเชิงเนื้อหา โดยมีค่า IOC ตั้งแต่ 0.6-1.0 ค่าความเที่ยง (KR-21) เท่ากับ 0.7 แบบทดสอบที่พัฒนาขึ้นมีคะแนนจุดตัดเท่ากับ .34 ในสเกลของคะแนนความสามารถ θ ซึ่งมีค่าเท่ากับ 534 คะแนนในช่วงสเกลของการประเมินนักเรียนระดับนานาชาติ (PISA)

2. ระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต ซึ่งประกอบด้วยเทคโนโลยีการจัดสอบและการทดสอบผ่านอินเทอร์เน็ต โดยมีองค์ประกอบ 4 องค์ประกอบ ได้แก่

2.1 ระบบการนำเข้าข้อสอบสู่ระบบ ระบบนี้เป็นระบบการนำเข้าข้อสอบ โดยต้องมีการพิมพ์ข้อสอบเข้าสู่ระบบรายข้อ ข้อสอบที่นำเข้าจะนำไปจัดเป็นแบบทดสอบเพื่อใช้เป็น

แบบทดสอบสำหรับการทดสอบ ข้อสอบที่นำเข้าเป็นได้ทั้งข้อความ ภาพ และเสียง เพื่อให้สามารถทำการวัดความสามารถทางวิทยาศาสตร์ของผู้สอบได้ครอบคลุม

2.2 ระบบการเขียน/ปรับข้อสอบ ระบบนี้เป็นระบบที่พัฒนาขึ้นมาเพื่อให้ผู้ใช้ระบบสามารถปรับปรุงแก้ไขข้อสอบให้ถูกต้อง และตรงตามความต้องการของผู้ใช้

2.3 ระบบการสอบผ่านอินเทอร์เน็ต ระบบการสอบเป็นระบบที่ประกอบด้วยข้อสอบและสื่อการนำเสนอ คือ การนำเสนอข้อสอบผ่านหน้าจอคอมพิวเตอร์ ผู้สอบจะต้องมีรหัสเข้าสอบเพื่อใช้จำแนกผลการสอบ และรายงานผลการสอบเป็นรายบุคคล ซึ่งผู้สอบสามารถเข้าสู่ระบบได้ทั้งคอมพิวเตอร์ และโทรศัพท์มือถือ

2.4 ระบบการรายงานผลการสอบ ระบบรายงานผลเป็นระบบที่วิเคราะห์ผลการทดสอบรายบุคคล โดยรายงานคะแนนสอบในรูปของคะแนนดิบ และผลการสอบผ่าน หรือไม่ผ่าน

3. ผลการตรวจสอบคุณภาพของระบบการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นผ่านอินเทอร์เน็ต

3.1 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning, DIF) ของแบบทดสอบทั้งสองแบบ พบว่ามีข้อสอบที่ทำหน้าที่ไม่ต่างกันจำนวน 26 ข้อ (ร้อยละ 96.30) มีข้อสอบ 1 ข้อที่ทำหน้าที่ต่างกัน (ร้อยละ 3.6) คือ ข้อที่ 13 ที่ทำหน้าที่ต่างกัน คือเป็นข้อที่มีความยากสูงสำหรับการทดสอบแบบ PBT แต่ไม่ค่อยยากสำหรับผู้สอบที่สอบแบบ IBT

3.2 ความคลาดเคลื่อนมาตรฐานของการวัด (standard error) สารสนเทศของแบบทดสอบ (information) ของแบบทสอบแบบ IBT และ PBT ไม่ต่างกัน สะท้อนให้เห็นว่าเป็นการวัดที่เท่าเทียมกัน โดยพิจารณาจากความคลาดเคลื่อนมาตรฐานของการวัด และสารสนเทศของแบบทดสอบทั้งสองแบบมีค่า และรูปแบบลักษณะใกล้เคียงกัน

3.3 คะแนนสอบของผู้สอบกลุ่ม IBT กับกลุ่ม PBT ไม่ต่างกัน โดยกลุ่ม IBT มีค่าเฉลี่ย 500.06, SD=100.09 กลุ่ม PBT มีค่าเฉลี่ย 499.94, SD=91.00 ซึ่งไม่ต่างกันอย่างมีนัยสำคัญทางสถิติ (t(305), =1.08, p=.28)

3.4 ส่วนเบี่ยงเบนของคะแนนสอบของกลุ่ม IBT และ PBT ใกล้เคียงกัน โดยส่วนเบี่ยงเบนมาตรฐานที่คำนวณจากคะแนนดิบของกลุ่ม IBT มีค่า 3.35 และ กลุ่ม PBT มีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.02 และถ้าหากพิจารณาค่าส่วนเบี่ยงเบนมาตรฐานจากคะแนนที่วิเคราะห์จากทฤษฎีการตอบข้อสอบ พบว่า กลุ่ม IBT และ PBT มี ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 100.09 และ 91.00 ตามลำดับ

4. การนำการประเมินการรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นที่พัฒนาขึ้นไปให้ครูและนักเรียนทั่วประเทศใช้ผ่านอินเทอร์เน็ต มีครูที่แจ้งความประสงค์ให้นักเรียนใช้ 46 คน โดยมีนักเรียนใช้ระบบจำนวน 675 คน ผลการสอบถามครู พบว่านักเรียนที่ใช้ระบบมีความสนใจในการทดสอบและครูเห็นว่าระบบที่พัฒนาขึ้นมีประโยชน์ต่อครูในด้านการประเมินระหว่างเรียน ครูสามารถใช้การ

ประเมินนี้เพื่อประเมินการเรียนรู้ของนักเรียน และการทดสอบผ่านคอมพิวเตอร์ช่วยให้นักเรียนมีความกระตือรือร้นในการทดสอบมากกว่าการสอบแบบทั่วไป นอกนี้การทดสอบผ่านคอมพิวเตอร์ยังช่วยพัฒนาทักษะการใช้คอมพิวเตอร์ของนักเรียนด้วย

อภิปรายผลการวิจัย

ในการวิจัยครั้งนี้ มีประเด็นที่ควรนำมาอภิปรายดังนี้

1. แบบทดสอบการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ตสำหรับนักเรียนมัธยมศึกษาตอนต้นมีความตรง และความเที่ยงสอดคล้องกับสมมติฐานที่ตั้งไว้ และสอดคล้องกับงานวิจัยของ Quellmalz, Davenport, Timms, DeBoer, Jordan, Huang, Buckley (2013) ที่พบว่าการทดสอบผ่านเทคโนโลยีสามารถวัดความรู้และทักษะขั้นสูงของผู้เรียนได้ โดยในการวิจัยครั้งนี้ความตรงพิจารณาจากความสัมพันธ์ของคะแนนสอบกับเจตคติ ซึ่งมีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ และสอดคล้องกับทฤษฎี และผลการสังเคราะห์งานวิจัยเกี่ยบกับสองตัวแปรนี้ (Wilson, 1983) ส่วนความเที่ยงในการวิจัยครั้งนี้พิจารณาจากสารสนเทศของแบบทดสอบ ซึ่งมีค่าสูงในระดับที่ยอมรับได้ (Hambleton, Swaminathan, & Rogers, 1991)

2. ระบบที่พัฒนาขึ้นมีประโยชน์ในการส่งเสริมการจัดการศึกษาที่ส่งเสริมการเรียนรู้ตลอดชีวิต กล่าวคือ การพัฒนาระบบประเมินการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ตเพื่อส่งเสริมการเรียนรู้วิทยาศาสตร์ขั้นสูงระดับมัธยมศึกษาตอนต้นในการวิจัยครั้งนี้ มุ่งพัฒนาระบบการประเมินโดยใช้เทคโนโลยีมาเป็นส่วนหนึ่งของการประเมินเพื่อให้สามารถพัฒนาระบบการประเมินที่ใช้ได้ง่าย และส่งเสริมการเรียนรู้ตลอดชีวิตของนักเรียน โดยระบบที่พัฒนาขึ้นจะช่วยให้ครู และนักเรียนสามารถเข้าถึงได้ตลอดเวลา ตลอดจนนักเรียนทุกกลุ่มไม่ว่าจะอยู่ภูมิภาคใดของโรงเรียนสามารถเข้าถึงระบบการประเมินได้อย่างเท่าเทียมกัน การพัฒนาดังกล่าวต้องมีความมั่นใจว่าระบบที่พัฒนาขึ้นจะต้องมีคุณภาพไม่น้อยกว่าระบบการประเมินที่ใช้กันอย่างทั่วไป คือ ระบบการประเมินที่ใช้การเขียนตอบ (paper-based testing)

3. ระบบที่พัฒนาขึ้นมีประสิทธิภาพในการวัดเทียบเคียงได้กับระบบการประเมินที่ใช้การเขียนตอบ (paper-based testing) ในการวัดผลการศึกษา การทดสอบหลายๆ รูปแบบที่เท่าเทียมกันควรวัดเนื้อหาเดียวกัน มีความเที่ยงใกล้เคียงกัน ค่าเฉลี่ยของคะแนนสอบใกล้เคียงกัน ความแปรปรวนของคะแนนสอบใกล้เคียงกัน (von Davier, & Holland, 2003) ซึ่งเป็นคุณสมบัติอย่างน้อยที่ควรรายงานเพื่อแสดงว่าแบบทดสอบมีคุณสมบัติทางการวัดใกล้เคียงกัน ในการวิจัยครั้งนี้ผู้วิจัยได้เปรียบเทียบ ความเที่ยง ค่าเฉลี่ยของคะแนนสอบ และความแปรปรวนของคะแนนสอบ ระหว่างกลุ่มผู้สอบ IBT และ PBT พบว่า ทั้งสองกลุ่มมีค่าสถิติใกล้เคียงกัน นอกจากนี้ การวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบทั้งสองประเภท พบว่า ข้อสอบส่วนใหญ่วัดได้เหมือนกันทั้งกลุ่ม IBT และ CBT สะท้อนให้เห็นว่า ระบบการทดสอบแบบ IBT และ PBT มีประสิทธิภาพในการวัดใกล้เคียงกัน (Holland & Wainer, 1993) และมีความเหมาะสมในการแปลผลคะแนน และการนำไปใช้ใกล้เคียงกัน (Messick, 1989, 1995) แต่ระบบการทดสอบ IBT ช่วยกระตุ้นความสนใจในการสอบ และเข้าถึงกลุ่มผู้สอบได้กว้างขวางมากกว่า

4. นอกเหนือจาก ระบบการทดสอบที่พัฒนาขึ้นมีคุณภาพในการด้านการวัดแล้ว ครูที่ใช้ระบบการทดสอบผ่านอินเทอร์เน็ตมีความเห็นว่า ระบบการทดสอบมีประโยชน์ในการประเมินและพัฒนาผู้เรียน ซึ่งจะให้ข้อมูลสำหรับการปรับปรุงการเรียนการสอนของครู และการปรับปรุงการเรียนของนักเรียน ดังนั้น จึงมีความเป็นไปได้ที่หากนำไปใช้ในระดับกว้าง จะทำให้เป็นการส่งเสริมการวัดและประเมินผลการเรียนเพื่อพัฒนาเรียน ตามแนวทางการประเมินเพื่อพัฒนา (assessment for learning) (McMillan, 2008)

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ควรมีการนำระบบการทดสอบผ่านอินเทอร์เน็ตไปใช้สนับสนุนการจัดการศึกษาของครูและโรงเรียน เพื่อเป็นการเพิ่มโอกาสทางการศึกษา และช่วยลดความเหลื่อมล้ำทางการศึกษา โดยอาจนำเนื้อหาวิชาทั้ง 8 กลุ่มสาระวิชา มาพัฒนาเป็นแบบทดสอบผ่านอินเทอร์เน็ตเพื่อให้นักเรียนและครูได้ใช้ในการจัดการเรียนการสอนที่เน้นการพัฒนาผู้เรียน (formative assessment)

2. ในการแปลความหมายคะแนน ไม่ควรเน้นการเปรียบเทียบระหว่างผู้เรียน แต่ควรใช้คะแนนเป็นส่วนหนึ่งของการวิเคราะห์จุดอ่อน จุดแข็งของผู้เรียน เพื่อนำไปสู่การได้ข้อมูลสำหรับการพัฒนาผู้เรียนให้มีความรอบรู้ตามหลักสูตร

ข้อเสนอแนะในการวิจัยครั้งต่อไป

ในการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเปรียบเทียบคุณสมบัติการวัดทางจิต คือ ความเที่ยง การทำหน้าที่ต่างกันของข้อสอบ ของการทดสอบทั้งสองแบบ ในการวิจัยครั้งต่อไปควรมีการศึกษาเพิ่มเติมในประเด็นที่ทันสมัย เช่น คุณสมบัติของการเทียบเคียงคะแนน (population invariance) (Dorans, & Holland, 2000) ซึ่งเป็นประเด็นของการศึกษาความสามารถในการเทียบเคียงกัน (comparability) ในมุมมองของการปรับเทียบคะแนน (test equating) ซึ่งผลการวิจัยเช่นนี้จะช่วยให้มีความมั่นใจได้ว่าระบบที่พัฒนาขึ้นมีประสิทธิภาพไม่แตกต่างจากระบบการทดสอบเขียนตอบที่ใช้กันอยู่ทั่วไป

ข้อจำกัดของการวิจัย

การวิจัยครั้งนี้มีข้อจำกัดสองประเด็นสำคัญ คือ จำนวนข้อสอบ และกลุ่มตัวอย่างที่ใช้ในการวิจัยค่อนข้างน้อย เนื่องจากนักเรียนบางกลุ่มไม่มีคอมพิวเตอร์ที่ต่อกับอินเทอร์เน็ตที่บ้าน การทดสอบจึงพยายามใช้ข้อสอบที่น้อย เพื่อให้นักเรียนสามารถทำการทดสอบได้เสร็จภายใน 1 คาบ ที่ครูสอนปกติ

ผลผลิต (output) ที่เกิดขึ้นในช่วงที่ได้รับทุน

1. บทความวิจัย

สังวรณ์ งัดกระโทก และ ชุติวัฒน์ สุวัตถิพงศ์ (2561) ความตรงและความเที่ยงของผลการประเมินการรู้วิทยาศาสตร์ผ่านระบบอินเทอร์เน็ต [อยู่ระหว่างการเผยแพร่]

สังวรณ์ งัดกระโทก และ ชุติวัฒน์ สุวัตถิพงศ์ (2561) ความเท่าเทียมกันของการวัดการรู้วิทยาศาสตร์: การเปรียบเทียบระหว่างรูปแบบการประเมินผ่านคอมพิวเตอร์ และรูปแบบการประเมินด้วยการเขียนตอบ [อยู่ระหว่างการเผยแพร่]

รายงานสรุปการเงิน ประจำปีงบประมาณ 2558

รหัสโครงการ (NRPM 13 หลัก) 25583300700265866

โครงการส่งเสริมการวิจัยในอุดมศึกษาและพัฒนามหาวิทยาลัยวิจัยแห่งชาติ

สำนักงานคณะกรรมการการอุดมศึกษา

ชื่อมหาวิทยาลัย มหาวิทยาลัยสุโขทัยธรรมาธิราช

ชื่อโครงการ การพัฒนาระบบประเมินการรู้วิทยาศาสตร์ผ่านอินเทอร์เน็ตเพื่อส่งเสริมการเรียนรู้วิทยาศาสตร์ขั้นสูงระดับมัธยมศึกษาตอนต้น

ชื่อหัวหน้าโครงการ ผู้ช่วยศาสตราจารย์ ดร. สังวรณ์ งัดกระโทก

รายงานในช่วงตั้งแต่วันที่ ถึงวันที่ 15 ธันวาคม 2560

ระยะเวลาดำเนินการ 2 จำนวน ปี 10 เดือน

รายจ่าย

หมวด	งบประมาณรวมทั้งโครงการ (บาท)	ค่าใช้จ่ายงวดปัจจุบัน (บาท)	คงเหลือ (บาท)
1. ค่าตอบแทน	60,000	60,000	-
2. ค่าจ้าง	100,000	100,000	-
3. ค่าวัสดุ	20,000	20,000	-
4. ค่าใช้สอย	70,000	70,000	-
5. ค่าใช้จ่ายอื่นๆ	-	-	-
รวม	250,000	250,000	-

จำนวนงบประมาณที่ได้รับ

- งวดที่ 1 จำนวน 150,000 บาท เมื่อ พ.ศ. 2558
- งวดที่ 2 จำนวน 100,000 บาท เมื่อ พ.ศ. 2560

รวม 250,000 บาท

ลงนามหัวหน้าโครงการ

วันที่ 15 ธันวาคม 2560

ลงนามเจ้าหน้าที่การเงินโครงการ

วันที่ 15 ธันวาคม 2560

บรรณานุกรม

ชินภัทร์ ภูมิรัตน. (2544). *การวิจัยเพื่อพัฒนานโยบายการปฏิรูปการศึกษาของไทย*. รายงานผลการวิจัย.

นงลักษณ์ วิรัชชัย. (2545). *กระบวนการปฏิรูปเพื่อพัฒนาคุณภาพการเรียนรู้: การประเมินและการประกัน*. สำนักงานเลขาธิการสภาการศึกษา .กรุงเทพมหานคร: ห้างหุ้นส่วนจำกัด วี.ที.ซี .คอมมิวนิเคชั่น.

สังวรณ์ งัดกระโทก. (2552). *คุณภาพการสอนวิทยาศาสตร์และความสามารถทางวิทยาศาสตร์ของนักเรียนไทย: ข้อค้นพบและข้อเสนอเชิงนโยบายจากการประเมินนักเรียนระดับนานาชาติ*. รายงานผลการวิจัย.

Adams, W. K., Reid, S., Lemaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., & Wieman, C. E. (2008). A study of educational simulations: Part 1—Engagement and learning. *Journal of interactive learning research, 19*, 397-419.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Virginia: Association for Supervision and Curriculum Development.

Buckley, B. C., Gobert, J., Horwitz, P., & O'Dwyer, L. (2010). Looking inside the black box: Assessing model-based learning and inquiry in BioLogica. *International Journal of Learning Technology, 5*, 166-190.

von Davier, A., & Holland, P. W. (2003). *The Kernel Method of Test Equating*. Springer: New York.

Dorans, N. J., & Holland, P .W. (2000). Population invariance and the equatability of tests: Basic theory and the Linear Case. *Journal of Educational measurement, 37(4)*, 281-306.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury, Park, CA: Sage.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Sage: New York.

Lau, K.-S. (2009). A critical examination of PISA's assessment on scientific literacy. *International journal of science and mathematics education, 7*, 1061-1088.

Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantages of technology to promote knowledge integration*. New York: NY: Routledge.

McMillan, J. H. (2008). Formative classroom assessment: The keys to improving student achievement. In J. McMillan (Ed.), *Formative classroom assessment* (pp. 1-7).New York: Teacher College.

Messick, S. (1989). Test validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13-103). New York: American Council on Education and Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50(9)*, 741-749.

Mislevy, R. J., & Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. Princeton, NJ: Educational Testing Service.

Perie, M., Marion, S., & Gong, B. (2007). A framework for considering interim assessments. National Center for the Improvement of Educational Assessment.

Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G. E., Jordan, K. A., Huang, C. W., & Buckley, B. C. (2013). Next generation environment for assessing and promoting complex science learning. *Journal of Educational Psychology, 105(4)*, 1100-1114.

Quellmalz, E. S., & Pelligrino, J. W. (2009). *Technology and testing science*, 323, 75-79.

Stiggins, R. J. (2008). Conquering the formative assessment frontier. In J. McMillan (Ed.), *Formative classroom assessment* (pp. 8-28).New York: Teacher College.

Willson, V. (1983). A meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college. *Journal of research in science teaching*. 20(9), 838-850.

ภาคผนวก

-

ประวัติผู้วิจัย

1. ผู้ช่วยศาสตราจารย์ ดร. สังวรณ์ งัดกระโทก เป็นอาจารย์ประจำสาขาวิชาศึกษาศาสตร์ แขนงวิชาการวัดและประเมินผลการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช จบการศึกษาระดับปริญญาเอกสาขา Measurement and Quantitative Methods จาก Michigan State University ประเทศสหรัฐอเมริกา มีความเชี่ยวชาญด้านการวิจัย สถิติ การออกแบบการวิจัย การวิจัยเชิงประเมิน การวัดและประเมินผลการศึกษา การประเมินการสอนวิทยาศาสตร์ อีเมลล์ติดต่อ sungworn@hotmail.com

2. อาจารย์ ดร. ชุติวัฒน์ สุวัตถิพงศ์ เป็นอาจารย์ประจำสำนักเทคโนโลยีการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช จบการศึกษาระดับปริญญาเอกสาขาเทคโนโลยีและสื่อสารการศึกษา จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มีความเชี่ยวชาญด้านคอมพิวเตอร์ และการออกแบบการเรียนรู้ผ่านเทคโนโลยี อีเมลล์ติดต่อ chu9372@hotmail.com