

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดและประเมินผลการศึกษา เป็นองค์ประกอบหนึ่งที่สำคัญยิ่งของการพัฒนาคุณภาพการศึกษาในทุกระดับ เพราะผลที่ได้นอกจากจะชี้ถึงความสำเร็จในการจัดการศึกษาแล้ว ยังเป็นข้อมูลพื้นฐานที่ใช้ประกอบการตัดสินใจของครูและผู้ที่เกี่ยวข้องในการวางแผนปรับปรุงวิธีการเรียนการสอน การแนะแนว การจัดหลักสูตร แบบเรียน การใช้อุปกรณ์การสอน ตลอดจนการจัดระบบบริหารทั่วไปของโรงเรียน จึงถือว่าเป็นความจำเป็นที่จะต้องได้ผลจากการวัดที่มีความถูกต้องแม่นยำ และมีคุณภาพให้มากที่สุดเท่าที่จะสามารถจัดกระทำได้ ดังนั้นการพัฒนาแบบสอบซึ่งเป็นเครื่องมือสำคัญอย่างหนึ่งซึ่งช่วยให้การวัดผลมีประสิทธิภาพ จึงเป็นสิ่งสำคัญและจำเป็นยิ่ง

การวัดและประเมินผลที่ใช้กันอยู่ในปัจจุบันมี 2 แบบ คือ การวัดผลแบบอิงกลุ่ม (Norm-Referenced Measurement) และการวัดผลแบบอิงเกณฑ์ (Criterion Referenced Measurement) การวัดผลแบบอิงกลุ่ม เป็นการวัดที่ เน้นความแตกต่างระหว่างบุคคลโดยนำความสามารถของแต่ละคนไป เปรียบเทียบกับคนในกลุ่มทั้งหมดที่สอบแบบสอบเดียวกัน (Popham 1978: 24-25) เพื่อจัดลำดับความสามารถของบุคคล ซึ่งไม่ค่อนำถึงระดับความรู้หรือความสามารถด้านเนื้อหาวิชาโดยตรง การวัดและประเมินผลแบบที่เหมาะสมที่จะใช้ตัดสินใจเลือกสรรคนหรือประเมินว่าใคร เก่งกว่ากัน ส่วนการวัดผลแบบอิง เกณฑ์ เป็นการวัดเพื่อมุ่งประเมินความสามารถของบุคคลว่าอยู่ตรงระดับใด เมื่อเทียบกับ เกณฑ์ที่กำหนดขึ้น (Popham 1978: 92-95) โดยมีวัตถุประสงค์สำคัญ เพื่อแยกผู้ เรียนรู้แล้วออกจากผู้ที่ยังไม่รู้ เพื่อสำรวจความก้าวหน้า ตลอดจนวินิจฉัยความสามารถทางการ เรียนรู้ของผู้เรียนแต่ละคน เป็นการมุ่งส่งเสริมความสามารถของผู้เรียนให้พัฒนาจนบรรลุเป้าหมายที่ต้องการได้ทุกคน อันเป็นลักษณะของการวัด เพื่อปรับปรุงและพัฒนา ซึ่งสอดคล้องกับแนวคิดของการจัดการศึกษาในปัจจุบัน

แบบสอบที่ใช้ในระบบการวัดผลแบบอิง เกณฑ์ สามารถจำแนกได้เป็น 2 ชนิดใหญ่ ๆ (บุญเชิด ภิญญอนันต์พงษ์ 2526: 29) ได้แก่

1. แบบสอบอิงจุดประสงค์ (Objective-Referenced Test)
2. แบบสอบอิงมวลความรู้หรืออิงโดเมน (Domain-Referenced Test)

แบบสอบอิงจุดประสงค์ หมายถึง แบบสอบที่บรรจุเนื้อหาสาระของข้อสอบตามจุดประสงค์เชิงพฤติกรรม ส่วนมากจะมีการกำหนดระดับเกณฑ์หรือมาตรฐาน เพื่อบ่งชี้ระดับความรู้ของผู้สอบ ผลที่ได้จากการสอบวัดจะนำไปเทียบกับจุดประสงค์การสอนว่าผู้เรียนเกิดการเรียนรู้ตามจุดประสงค์การสอนหรือไม่ (บุญเชิด ภิญโญนันต์พงษ์ 2526: 6-29)

แบบสอบอิงโดเมน หมายถึง แบบสอบที่ประกอบด้วยกลุ่มตัวอย่างของข้อสอบซึ่งอาจได้มาจากการสุ่มอย่างง่าย (Simple Random Sampling) หรือสุ่มแยกประเภท (Stratified Sampling) จากประชากรข้อสอบทั้งหมดที่ได้รับการกำหนดขึ้นตามโดเมนที่อ้างอิงถึง (Millman 1974: 17-21) ผลที่ได้จากการสอบวัดจะนำไปใช้ประมาณค่าความสามารถของผู้สอบที่จะสามารถตอบข้อสอบถูกใน เซทหรือจักรวาลของงานที่กำหนดให้ทำ (Domain) ซึ่งได้นิยามขอบเขตไว้อย่างชัดเจน (Hively 1974: 5-10) กล่าวคือ ในการแปลความหมายของคะแนนสอบ จะใช้จำนวนข้อที่ตอบถูกในกลุ่มตัวอย่างของโดเมนนี้มาคำนวณหาค่าอัตราส่วนที่แท้จริง เมื่อเทียบกับพฤติกรรมทั้งหมดในโดเมน

เบเกอร์ (Baker 1974: 12-16) ได้ชี้ถึงข้อแตกต่างของการพัฒนาข้อสอบอิงโดเมนและข้อสอบอิงจุดประสงค์โดยทั่วไป ว่าเป็น ข้อสอบอิงวัตถุประสงค์โดยทั่วไป จะเน้นคุณลักษณะของการเรียนการสอนตามที่เบเกอร์ (Mager 1962) ได้กำหนดไว้ว่าประกอบด้วยวัตถุประสงค์ที่ระบุเป็นพฤติกรรม เกณฑ์การให้คะแนน สภาพการณ์เงื่อนไขของการสอบ และตัวอย่างข้อสอบซึ่งอาจมีไว้ให้ สำหรับข้อสอบอิงโดเมนจะพัฒนาขึ้นจากคุณลักษณะต่าง ๆ ของข้อสอบอิงวัตถุประสงค์ดังกล่าวแล้ว และจะให้ความสำคัญกับสิ่งต่าง ๆ เพิ่มเติมอีกคือ คำอธิบายโดเมน (Domain Description) ขอบเขตเนื้อหา (Content Limits) เกณฑ์หรือโดเมนของตัวลวง (Criterion or Distractor Domain) รูปแบบข้อสอบ (Format) คำชี้แจง (Direction) และ ตัวอย่างข้อสอบ (Sample Items)

การประเมินผลสัมฤทธิ์ทางการเรียนในปัจจุบัน มีความต้องการกำหนดเกณฑ์ความสามารถจริงของผู้เรียนอย่างถูกต้อง ดังนั้นการสร้างแบบสอบจึงจำเป็นต้องวางแผนการสร้างให้มีรายละเอียดต่าง ๆ มากขึ้น ซึ่งฟวงรัตน์ ทวีรัตน์ (2529: 69) และโกวิท ประวาลพฤกษ์ (2523: 36-48) ได้เสนอความเห็นที่สอดคล้องกับแนวคิดของ แฮริส และสจิวค (Harris and Stewart 1971: 1) ว่าแบบสอบอิง เกณฑ์ที่แท้จริงควรเป็นแบบสอบอิงโดเมน

ในปัจจุบัน กรมวิชาการ กระทรวงศึกษาธิการได้นำหลักการของการวัดและประเมินผลแบบอิง เกณฑ์บรรจุ เข้าไว้ในระเบียบว่าด้วยการประเมินผลการเรียนระดับมัธยมศึกษาตอนต้น พุทธศักราช 2521 พ.ศ. 2529 ซึ่งได้กำหนดไว้ว่า ให้สถานศึกษาประเมินผลการเรียน ให้สอดคล้องกับจุดประสงค์การเรียนรู้อันหนึ่งของแต่ละรายวิชา ผลการประเมินให้นำมาใช้เพื่อปรับปรุง การเรียนการสอน และเพื่อตัดสินผลการเรียน ถ้าผลการประเมินของผู้เรียนไม่ผ่านจุดประสงค์ การเรียนรู้ให้ดำเนินการสอนซ่อมเสริม เพื่อพัฒนาผู้เรียนให้เกิดการเรียนรู้ต่อไป (กรมวิชาการ 2529: 1) ซึ่งจะเห็นว่าเป็นการประเมินผลแบบอิง เกณฑ์ในลักษณะอิงวัตถุประสงค์การเรียนรู้ หรือจุดประสงค์เชิงพฤติกรรมนั่นเอง ซึ่งพิศิษฐ ตัณฑวนิช (2529: 58) ได้ให้ข้อสังเกตว่า แบบสอบอิง เกณฑ์ที่พัฒนาขึ้นใช้ในประเทศไทย โดยเฉพาะในระดับโรงเรียนยังยึดวัตถุประสงค์ เฉพาะ เป็นหลักในการสร้างแต่ไม่ได้มีการกำหนดขอบเขต เนื้อหาให้เด่นชัด หรือไม่มีกรอบ ขอบเขตใด เมินที่ผู้สร้างใช้เป็นหลักในการเขียนข้อสอบ และในการนำแบบสอบอิงจุดประสงค์ ไปใช้ปฏิบัติจริงก็ประสบปัญหาหลายอย่าง (สงบ ลักษณะ 2525: 1-14) เช่น รายวิชาหนึ่ง ๆ มีจุดประสงค์มากมาย ปัญหาที่เกิดกับครูผู้สอนก็คือ การมองไม่เห็นบูรณาการหรือการโยงความ สัมพันธ์ของจุดประสงค์เหล่านี้ ไม่ทราบว่าจุดประสงค์ใด เชื่อมกับจุดประสงค์ใด จุดประสงค์ ใดสำคัญมากสำคัญน้อย ถ้าผู้เรียนทำข้อสอบบางจุดประสงค์ได้แต่ทำไม่ได้ในบางจุดประสงค์ จะ สรุปลผลและตัดสินอย่างไร ทำให้มองไม่เห็นภาพจน์ของผลสัมฤทธิ์ทางการเรียนของผู้เรียนว่า เค้นด้วยประการใดภายใต้เนื้อหาย่อย ๆ ของแต่ละรายวิชา และสมศักดิ์ ลินตุระ เวชญ์ (2525: 59) ได้กล่าวถึงจุดอ่อนของแบบสอบอิงจุดประสงค์ซึ่งสอดคล้องกับข้อคิดเห็นของไฮเฟล และคณะ (Hively, et al. 1968: 275-290) ว่า ในการสร้างแบบสอบถ้ากำหนดจุดประสงค์ไม่ชัดเจนหรือไม่ดีแล้ว ก็จะทำให้การวัดไม่ถูกจุด หรือแคบเกินไป นอกจากนี้การสอนแบบอิงจุด ประสงค์ยังไม่ให้ความหมายในเชิงอ้างอิงไปสู่สัดส่วนของ เนื้อหาทั้งหมดในเรื่องนั้นแต่อย่างใด ซึ่งรอยด์ และฮาลาดินา (Roid and Haladyna 1982 อ้างถึงใน สมหวัง พิธิยานุวัฒน์ 2529: 22) กล่าวว่าในการกำหนดคะแนน เกณฑ์ในการทดสอบอิง เกณฑ์ เป็นสิ่งปราศจากความ หมาย ถ้าไม่มีการกำหนดการแปลความหมายของคะแนนใน เชิงปริ เชตหรือใด เมิน

การประเมินผล เพื่อพัฒนาการเรียนการสอนนั้น ควรทำตลอดเวลาในระหว่างการสอน และการประเมินผลที่ตีนั้น เมื่อทราบผลของผู้เรียนแล้วต้องบอกได้ว่า ผู้เรียนมีความสามารถเด่น หรือด้อยในเรื่องอะไร ตอนใด เพื่อผู้สอนจะได้ปรับปรุงแก้ไขผู้เรียนได้ถูกจุด ตลอดจนได้

ปรับปรุงวิธีการสอนของคนให้เหมาะสมด้วยการสอบวัดที่สามารถแปลผล เพื่อบ่งบอกความสามารถของผู้เรียนในลักษณะมวลความรู้ตามกลุ่มพฤติกรรมหรือขอบเขตเนื้อหาเป็นหน่วยย่อย ๆ โดยใช้แบบสอบอิงโคเมน เป็นวิธีการที่ดีที่สุด (โกวิท ประวาลพุกษ์ 2523: 36-48) ซึ่งสอดคล้องกับที่มิลแมน (Millman 1974) ได้กล่าวไว้ว่า แบบสอบอิงเกณฑ์ในรูปของแบบสอบอิงโคเมนสามารถบรรยายถึงการกระทำได้ดีกว่า

จุดเด่นของแบบสอบอิงโคเมนคือ สามารถช่วยจัดกลุ่มการกระจายกระจายของจุดประสงค์เข้าเป็นกลุ่มก้อน ช่วยในการแปลผลของการวัดที่มีความหมายยิ่งขึ้น กล่าวคือ คะแนนที่ได้สามารถให้ความหมายโดยตรงเกี่ยวกับกลุ่มงานที่กระทำได้ (Performance Tasks) ของผู้สอบ มีลักษณะบูรณาการ เป็นกลุ่มก้อน ได้ประเด็นสาระที่มีความหมาย สามารถสรุปประเด็นหลักที่ผู้เรียน เกิดการรอบรู้และที่ยังบกพร่อง ซึ่งเป็นข้อมูลที่มีประโยชน์ต่อการสอนซ่อมเสริม หรือต่อการพัฒนาการเรียนการสอนตามแนวทางของการใช้หลักสูตรใหม่ในปัจจุบัน (สงขล 2525: 1-14 และ เขียวดี วิบูลย์ศรี 2528: 34) ซึ่งการวัดโดยใช้แบบสอบอิงโคเมนมีความเหมาะสมมากที่สุด เมื่อเนื้อหาที่จะวัด เป็นโคเมนที่สามารถให้นิยามได้ชัดเจน จำนวนเนื้อหา มีขอบเขตจำกัด และกรอบของกลุ่มตัวอย่าง (Sampling Frame) มีรายการเนื้อหาทั้งหมดของโคเมนปรากฏอยู่หรือสามารถสร้างได้ครบถ้วน (Shaycoft 1979: 4)

อนึ่งในการสร้างและพัฒนาแบบสอบนั้น การวิเคราะห์ข้อสอบ (Item Analysis) จัดว่าเป็นสิ่งสำคัญและจำเป็นยิ่ง เพราะผลจากการวิเคราะห์นอกจากจะใช้เป็นแนวทางในการปรับปรุงข้อสอบให้ดีขึ้นแล้ว ยังช่วยให้ข้อมูลแก่ครูผู้สอนว่าจุดบกพร่องในการเรียนการสอนอยู่ที่ใด การวิเคราะห์ข้อสอบโดยอาศัยแนวคิดของทฤษฎีคลาสสิกอล (Classical Test Theory) ที่ใช้กันมาก่อนจนถึงปัจจุบันนั้น เป็นการมุ่งตรวจสอบคุณภาพแบบสอบโดยการวิเคราะห์ข้อสอบรายข้อ (Item Analysis) เพื่อพิจารณาค่าความยากของข้อสอบ (Item Difficulty) และค่าอำนาจจำแนก (Discriminating Power) และตรวจสอบคุณภาพแบบสอบทั้งฉบับโดยพิจารณาจากค่าความเที่ยง Reliability) และความตรง (Validity) ซึ่งในปัจจุบันการวิเคราะห์ข้อสอบตามแนวคิดของทฤษฎีคลาสสิกอลนี้ได้รับการวิจารณ์ว่ามีข้อจำกัดหลายประการ เช่น ค่าสถิติของข้อสอบ ได้แก่ ค่าความยาก และค่าอำนาจจำแนกของข้อสอบขึ้นอยู่กับกลุ่มตัวอย่างผู้สอบ ดังนั้นค่าสถิติของข้อสอบดังกล่าวจึงมีประโยชน์ในการสร้างแบบสอบเฉพาะ

กลุ่มที่มีลักษณะคล้ายคลึงกับกลุ่มที่ใช้ในการหาค่าสถิติของข้อสอบ เท่านั้น สำหรับค่าความเที่ยงของแบบสอบที่ได้ก็นิยามในลักษณะของแบบสอบคู่ขนาน ซึ่งมีโน้ตส์ของการวัดแบบเดียวกันที่จะสัมฤทธิ์ผลได้ในทางปฏิบัติ เพราะผู้สอบแต่ละคนมักจะได้ผลการสอบที่ต่างจากการสอบครั้งแรก ซึ่งอาจเกิดจากการลืม การพัฒนาทักษะใหม่ ๆ การเปลี่ยนแปลงของระดับแรงจูงใจหรือความวิตกกังวล เป็นต้น นอกจากนี้แนวคิดของทฤษฎีคลาสสิกอลยังมีจุดอ่อนอื่น ๆ อีก เช่น การเปรียบเทียบความสามารถระหว่างบุคคลทำได้ในวงจำกัด เฉพาะ เป็นผลการวัดด้วยข้อสอบชุดเดียวกันหรือคู่ขนานกัน เท่านั้น . และข้อตกลงเบื้องต้นที่ว่าความแปรปรวนของความคลาดเคลื่อนในการวัดมีค่าเท่ากันสำหรับผู้สอบทุกคน ซึ่งไม่สอดคล้องกับสภาพความเป็นจริง เพราะค่าความแปรปรวนของความคลาดเคลื่อนในการวัดจะแปรเปลี่ยนไปตามความสามารถของแต่ละบุคคล นอกจากนี้ยังไม่มีสารสนเทศ (Information) ที่เป็นพื้นฐานที่จะช่วยทำนายว่าผู้สอบคนหนึ่งจะสามารถทำข้อสอบได้ดีเพียงใด เมื่อได้เผชิญกับข้อสอบข้อหนึ่ง ๆ ซึ่งควรจะมีการประมาณค่าความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้อง ด้วยค่าที่เหมาะสมตรงกับระดับความสามารถของผู้สอบ อันเป็นสารสนเทศที่จำเป็น (Hambleton and Swaminathan 1985: 1-3)

จากจุดอ่อนดังกล่าว ทำให้นักวัดผลจำนวนมากหันมาสนใจและพัฒนาทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT หรือ Latent Trait Theory หรือ Item Characteristic Curve Theory หรือ Modern Test Theory) และสามารถนำมาใช้ในการวิเคราะห์เพื่อหาคุณภาพ เป็นรายข้อและทั้งฉบับได้ ซึ่งแฮมเบิลตัน (Hambleton 1979: 16-22) กล่าวว่า ถ้านำทฤษฎีนี้มาใช้ตรงตามข้อตกลงเบื้องต้น จะสามารถแก้ไขจุดอ่อนของทฤษฎีคลาสสิกอลได้ ซึ่งปัจจุบันก็มีการนำเอาแนวคิดของ IRT ไปใช้อย่างแพร่หลาย

IRT เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกับระดับความสามารถที่มีอยู่ภายในตัวผู้สอบ ซึ่งความสัมพันธ์ดังกล่าวเพื่ออธิบายได้โดยแสดงในลักษณะ โมเดลทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชันทางคณิตศาสตร์ ซึ่งเรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือโค้งลักษณะข้อสอบ (Item Characteristic Curve) โมเดลทางคณิตศาสตร์ที่นำมาใช้อธิบาย IRT มีหลายรูปแบบแต่ที่ได้รับการพัฒนาและมีการนำไปใช้กันมากที่สุดในหลาย ๆ สถานการณ์ของการสอบ เนื่องจากสะดวกในการใช้ คือ โมเดลโลจิสติก (Logistic Model) ซึ่งแบ่งเป็น 4 โมเดลย่อยตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบ และจากการศึกษาของยูริ (Urry 1970 อ้างถึงใน

สุพัทธ์ สุกมลสันต์ 2530: 57) พบว่าโมเดลโลจิสติกชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Model) เหมาะสำหรับแบบสอบเลือกตอบมากที่สุด เพราะมีการพิจารณาถึงค่าอำนาจจำแนก ค่าความยาก และค่าการเดา ในการอธิบายถึงความน่าจะเป็นที่ผู้สอบจะสามารถตอบข้อสอบได้ถูกต้อง ซึ่งสอดคล้องกับลักษณะและสถานการณ์ของการสอบที่เป็นจริงมากที่สุด

แนวคิดของทฤษฎีการตอบสนองข้อสอบในการตรวจสอบคุณภาพของแบบสอบ ทำได้โดยพิจารณาจากค่าฟังก์ชันสารสนเทศ (Information Function) ซึ่งถือว่าเป็นดัชนีที่ใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำ ค่าฟังก์ชันสารสนเทศนี้เป็นฟังก์ชันของค่าพารามิเตอร์ของข้อสอบที่เป็นดัชนีผสมโดยสร้างจากดัชนีที่กำหนดคุณภาพของข้อสอบหลาย ๆ ดัชนีด้วยกันจึงไม่ต้องพิจารณาคูณลักษณะของข้อสอบโดยแยกประเด็นดังเช่น แนวคิดของทฤษฎีคลาสสิกอล ค่าฟังก์ชันสารสนเทศที่ใช้พิจารณาคุณภาพของแบบสอบได้ใช้วิธีการพิจารณาจากค่าฟังก์ชันสารสนเทศแบบสอบ (Test Information Function) ซึ่งหาได้จากผลรวมของค่าฟังก์ชันสารสนเทศข้อสอบ (Item Information Function) โดยกำหนดว่า ณะระดับความสามารถใดที่ค่าฟังก์ชันสารสนเทศแบบสอบมีค่าสูง แสดงว่าแบบสอบมีคุณภาพที่ดีในการประมาณค่าความสามารถที่ระดับนั้น

จุดเด่นที่สำคัญประการหนึ่งของทฤษฎีการตอบสนองข้อสอบคือ ค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนข้ามกลุ่มย่อย ๆ ของผู้สอบที่เลือกมาจากกลุ่มประชากรกลุ่มเดียวกัน กล่าวคือ ภายในกลุ่มประชากรกลุ่มใหญ่ที่เราตั้งใจจะสร้างข้อสอบขึ้นมาใช้นั้น ไม่ว่าเราจะเลือกกลุ่มตัวอย่างกลุ่มย่อยกลุ่มใดมา เพื่อใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ ค่าประมาณของพารามิเตอร์เหล่านั้นจะมีค่าคงที่ ซึ่งคุณลักษณะ เช่นนี้ เป็นประโยชน์มากในการสร้างและพัฒนาข้อสอบ (ผจงจิต อินทสุวรรณ 2525: 67) โดยเฉพาะแบบสอบอิงโดเมน ซึ่งจะต้องสุ่มข้อสอบมาจากโดเมนที่มีข้อสอบในลักษณะเดียวกันจำนวนมาก ๆ เช่น สามารถวัดคุณลักษณะเดียวกัน และมีระดับความยากเท่า ๆ กัน... ซึ่งการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ จึงนับว่าช่วยแก้ปัญหานี้ได้ เพราะจะได้ค่าพารามิเตอร์ความยากคงที่ เนื่องจากเป็นค่าที่เป็นอิสระจากกลุ่ม (สุพัทธ์ สุกมลสันต์ 2529: 18-19) นอกจากนี้ยังช่วยแก้ปัญหาของการใช้สถิติตามแนวคลาสสิกอลในแง่ที่ว่าค่าความยากและคะแนนโดเมน (Domain Score) อยู่กันคนละมาตรของการวัด ไม่อาจนำมาเทียบกันได้ ทั้งนี้เพราะการอ้างอิงค่าความยากของข้อสอบจะใช้

ผู้เข้าสอบทั้งหมด แต่การอ้างอิงคะแนนโดเมนจะใช้โดเมนของเนื้อหาที่พิจารณา (ดำรง
ศิริ เจริญ 2527: 12-21)

สำหรับคะแนนโดเมน (π) จะเป็นค่าสถิติที่ใช้คาดคะเนระดับผลสัมฤทธิ์ทางการเรียน
ที่ปราศจากอคติและพหุเพียงได้ ก็ต่อเมื่อมีการกำหนดกลุ่มของข้อสอบที่เป็นไปได้ในแต่ละโดเมน
ไว้ล่วงหน้า และข้อสอบในแบบสอบอิงโดเมนต้องเป็นกลุ่มตัวอย่างแบบสุ่มของข้อสอบในโดเมนนั้น
จุดมุ่งหมายอันหนึ่งของแบบสอบอิงโดเมนก็เพื่อวัดผลสัมฤทธิ์การประมาณค่าคะแนนโดเมนได้อย่าง
น่าเชื่อถือ ซึ่งจากแนวคิดของทฤษฎีตอบสนองข้อสอบนั้น เมื่อทราบค่าพารามิเตอร์ของข้อสอบ
หลายข้อที่อยู่ในโดเมนเดียวกัน หรือทราบค่าพารามิเตอร์ของข้อสอบชุดหนึ่งที่เป็นตัวแทนที่ดีของ
โดเมน จะสามารถหาความสัมพันธ์ระหว่างคะแนนโดเมน และคะแนนความสามารถที่แท้จริง
(θ) ได้เนื่องจากคะแนนโดเมน คือค่าคาดหวัง (Expected Value) ของคะแนนจากข้อสอบ
ทุกข้อในโดเมน (Hambleton and De Gruijter 1983: 358)

การวัดผลอิงโดเมน ยังนับว่าเป็น เรื่องใหม่ในวงการวัดผลสัมฤทธิ์ทางการศึกษาของ
ไทย โดยเฉพาะในระดับปฏิบัติในโรงเรียน ความรู้ในเรื่องของการวัดผลอิงโดเมน การสร้าง
เครื่องมือ การวิเคราะห์คุณภาพ ตลอดจนการแปลความหมายคะแนน ยังไม่แพร่หลายและไม่ได้
รับการพัฒนาย่างจริงจัง ซึ่งสมหวัง พิธิยานุวัฒน์ (2529: 18) กล่าวว่า "ผู้เขียนหวังว่า
ในอนาคตอันใกล้คงจะมีการพัฒนาโปรแกรมการทดสอบอิงปริเขตในระดับกลุ่มโรงเรียนและใน
ระดับเขตการศึกษาในประเทศไทย อันจะช่วยให้การสอนอย่างเป็นระบบในประเทศไทยประสบ
ความสำเร็จมากยิ่งขึ้น อันเป็นเป้าหมายสำคัญของการทดสอบที่มุ่งปรับปรุงพัฒนาและยกคุณภาพ
ของการศึกษาโดยส่วนรวมนั้นเอง..."

ในการสร้างและพัฒนาแบบสอบอิงโดเมนครั้งนี้ ผู้วิจัยสร้างแบบสอบอิงโดเมนวิชา
คณิตศาสตร์ เนื่องจากวิชาคณิตศาสตร์เป็นวิชาที่มีเนื้อหาเน้นทางด้านทักษะและสามารถแจกแจง
เป็นพฤติกรรมย่อย ๆ ได้ครบทุกลักษณะ คือมีส่วนประกอบของพฤติกรรมที่จำกัด (Limited
Number of Abilities) อันเป็นลักษณะที่เหมาะสมกับการสร้างแบบสอบอิงโดเมน และ
นอกจากนี้วิชาคณิตศาสตร์ยังจัดว่าเป็นวิชาที่มีความสำคัญมากทั้งในด้านการพัฒนาความคิดของ
ผู้เรียนและเป็นเครื่องมือในการนำไปสู่ความก้าวหน้าในวิทยาการด้านต่าง ๆ ตลอดจนมีความ
เกี่ยวข้องกับชีวิตประจำวันของบุคคล และในปัจจุบันกระทรวงศึกษาธิการได้กำหนดให้วิชาคณิต
ศาสตร์ เป็นรายวิชาบังคับและเลือกในระดับมัธยมศึกษาตอนต้นด้วย แต่ถึงแม้ว่าวิชาคณิตศาสตร์

จะมีความสำคัญมากดังกล่าว แต่ก็เป็นที่ทราบโดยทั่วไปว่า วิชาคณิตศาสตร์เป็นวิชาที่นักเรียนส่วนใหญ่ไม่ชอบและมีผลสัมฤทธิ์ทางการเรียนค่อนข้างต่ำ ยังต้องปรับปรุง (กรมวิชาการ 2528: 6) ซึ่งสาเหตุอาจมาจากตัวนักเรียนเอง วิธีการจัดการเรียนการสอนของครู เป็นต้น ซึ่งการพัฒนาข้อสอบอิงโดเมนวิชาคณิตศาสตร์ นอกจากจะนำไปสู่การสร้างเครื่องมือที่มีประสิทธิภาพแล้ว ความละเอียดของโดเมนจะช่วยให้ผู้เรียน ครู และผู้เกี่ยวข้องทราบถึงวัตถุประสงค์ เป้าหมาย กระบวนการเรียนการสอน ความก้าวหน้าในการเรียนและระดับความสามารถของผู้เรียนว่าสอดคล้องกับวัตถุประสงค์ปลายทางของการเรียนรู้มากน้อยเพียงใด โดยที่ข้อมูลดังกล่าวจะเป็นประโยชน์ในการประเมินผลการเรียนการสอนและการจัดสอนซ่อมเสริม ซึ่งจะก่อให้เกิดแรงจูงใจในการเรียน อันจะส่งผลถึงคุณภาพการเรียนการสอนวิชาคณิตศาสตร์ให้ดีขึ้นได้

ด้วยเหตุผลต่าง ๆ ดังกล่าวแล้ว ผู้วิจัยจึงสนใจที่จะสร้างและพัฒนาแบบสอบอิงโดเมนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 เรื่องพหุนาม ซึ่งเหตุผลในการที่เลือกศึกษาในระดับชั้นมัธยมศึกษาปีที่ 3 เนื่องจากผู้วิจัยเคยสอนวิชาคณิตศาสตร์ในระดับชั้นนี้มาก่อน มีความชำนาญในเนื้อหาพอสมควร ตลอดจนได้รู้ถึงปัญหาในการเรียนการสอน ซึ่งเป็นประโยชน์ในการสร้างแบบสอบอิงโดเมนนี้ โดยเฉพาะในแง่ของการวิเคราะห์เนื้อหา ความพลิกแพลงในการกำหนดตัวเลือก ตัวลวง เป็นต้น และในการสร้างแบบสอบได้อาศัยเทคนิคการเขียนข้อสอบจากการกำหนดลักษณะ เฉพาะของข้อสอบ (Item Specification) ตามวิธีของสงบ ลักษณะ (2525) ซึ่งเป็นการแปลงพฤติกรรมย่อยให้มีความเฉพาะเจาะจงวิธีหนึ่ง เป็นวิธีการที่พัฒนามาจากปัญหาของจุดประสงค์เชิงพฤติกรรมที่ไม่สามารถนิยามมวลความรู้ได้ชัดเจน ซึ่งเชื่อว่าจะช่วยให้ได้ข้อสอบที่มีความเป็นตัวแทนของขอบเขตเนื้อหาที่กำหนดขึ้น และนำผลสอบมาวิเคราะห์ข้อสอบรายข้อโดยใช้แบบจำลองโลจิสติก 3 พารามิเตอร์ เพื่อให้ได้แบบสอบอิงโดเมนวิชาคณิตศาสตร์ที่มีคุณภาพในการนำไปใช้ในการเรียนการสอน และเป็นแนวทางในการพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนในรายวิชาอื่นและระดับอื่น ๆ ต่อไป

วัตถุประสงค์ของการวิจัย

จุดประสงค์ของการวิจัยครั้งนี้ เพื่อประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) โมเดลโลจิสติก 3 พารามิเตอร์ ในการสร้างและพัฒนาแบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้น โดยมีวัตถุประสงค์เฉพาะคือ

1. เพื่อสร้างแบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่อง พหุนาม (Polynomial)
2. เพื่อใช้โมเดลโลจิสติก 3 พารามิเตอร์ในการวิเคราะห์แบบสอบอิงโดเมน ค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) และคัดเลือกข้อสอบตามเกณฑ์ที่เหมาะสมของค่าพารามิเตอร์
3. เพื่อตรวจสอบคุณภาพของแบบสอบที่สร้างขึ้นตามรูปแบบอิงโดเมน และรูปแบบของ IRT
4. เพื่อใช้โมเดลโลจิสติก 3 พารามิเตอร์กำหนดคะแนน เกณฑ์ของแบบสอบที่ได้ จากวิธีของแองกอฟ
5. เพื่อประมาณค่าความสามารถ (θ) และคะแนนโดเมน (π) ของผู้สอบ ตามรูปแบบของโมเดลโลจิสติก 3 พารามิเตอร์

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2531 ของโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในจังหวัดศรีสะเกษ เท่านั้น
2. เนื้อหาวิชาที่ใช้ในการสร้างแบบสอบอิงโดเมนครั้งนี้คือ เนื้อหาวิชาคณิตศาสตร์ เรื่องพหุนาม ชั้นมัธยมศึกษาปีที่ 3 (ค311) ตามหลักสูตรระดับมัธยมศึกษาตอนต้นพุทธศักราช 2521 ของกรมวิชาการ กระทรวงศึกษาธิการ

ข้อตกลงเบื้องต้น

1. นักเรียนทุกคนทำแบบสอบด้วยความสามารถ และตั้งใจจริง
2. เพศ อายุ และวุฒิของอาจารย์ผู้สอนที่ต่างกัน ไม่มีผลต่อผลการเรียนรู้ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนต่าง ๆ ที่ใช้เป็นกลุ่มตัวอย่าง

คำจำกัดความที่ใช้ในการวิจัย

1. แบบสอบอิงโดเมน (Domain Referenced Test) หมายถึง แบบสอบวัดผลสัมฤทธิ์ทางการเรียน ซึ่งมีลักษณะดังนี้
 - 1.1 วัดผลสัมฤทธิ์ของเนื้อหาและพฤติกรรมที่กำหนดขึ้นในขอบเขตที่ชัดเจน โดยคำนึงถึงความเป็นเอกพันธ์ของเนื้อหา ลักษณะพฤติกรรม และความยาวที่เหมาะสม
 - 1.2 เป็นตัวแทนของประชากรข้อสอบทั้งหมดที่วัดตามเนื้อหาและพฤติกรรมที่กำหนด
 - 1.3 การสร้างแบบสอบ สร้างขึ้นตามลักษณะเฉพาะของข้อสอบ (Item Specification) การกำหนดลักษณะเฉพาะของข้อสอบ เปรียบเสมือนการกำหนดขอบเขตประชากรข้อสอบ ข้อสอบที่ผลิตขึ้นจากลักษณะเฉพาะของข้อสอบ เปรียบเสมือนตัวแทนข้อสอบที่สุ่มมาจากประชากรของข้อสอบ (Hively 1974: 5)
2. ลักษณะเฉพาะของข้อสอบ หมายถึง แบบฟอร์มการกำหนดระเบียบหรือกฎเกณฑ์สำหรับเขียนข้อสอบให้รัดกุม รอบคอบชัดเจน สมบูรณ์ด้วยคุณภาพต่าง ๆ เช่น มีเหตุผล เป็นปรนัย วัดในสิ่งที่ต้องการวัดได้
3. โดเมน (Domain) หรือปริเขต หรือมวลความรู้ หมายถึง มวลพฤติกรรมหรือประชากรพฤติกรรมด้านพุทธิพิสัย ในขอบเขตเนื้อหาวิชาคณิตศาสตร์ เรื่องพหุนามซึ่งได้จากการนำเนื้อหาและพฤติกรรมมาผนวกกัน
4. สับโดเมน (Sub Domain) หมายถึง กลุ่มย่อย ๆ ของแต่ละโดเมน ซึ่งใช้ในการกำหนดลักษณะเฉพาะของข้อสอบของแบบสอบที่สร้างขึ้น
5. คะแนนโดเมน (Domain Score) หมายถึง ค่าคาดหวังของคะแนน (Expected Score) จากข้อสอบทุกข้อในโดเมน คำนวณจากค่าเฉลี่ยของผลรวมของโอกาสตอบถูกรายข้อของข้อสอบในแบบสอบในแต่ละค่าความสามารถ (๑)

6. คะแนนเกณฑ์ (Criterion: θ^*) หมายถึง ระดับความสามารถที่ใช้เป็นเกณฑ์มาตรฐานขั้นต่ำสุดที่จะตัดสินว่า เป็นผู้รอบรู้ และไม่รอบรู้

7. นักเรียน หมายถึง นักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2531 ของโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในจังหวัดศรีสะเกษ

8. โมเดลโลจิสติก 3 พารามิเตอร์ (Three-Parameter Logistic Model) หมายถึง รูปแบบหนึ่งของทฤษฎีการตอบสนองข้อสอบ ซึ่งคำนวณค่าพารามิเตอร์ของข้อสอบ ได้แก่ค่า a, b, c

9. ค่าความยาก (Item Difficulty: b) หมายถึง ดัชนีของข้อสอบที่แสดงถึงระดับความยากง่ายของข้อสอบ และแสดงถึงระดับความสามารถของบุคคลที่มีความน่าจะเป็นของการตอบข้อสอบข้อนั้นถูก $= (1+c)/2$ ในทางปฏิบัติ จะมีค่าระหว่าง -3 ถึง $+3$ ถ้า $b = -3$ แสดงว่า ข้อสอบง่ายมาก ถ้า $b = +3$ แสดงว่า ข้อสอบยากมาก

10. ค่าอำนาจจำแนก (Item Discriminating: a) หมายถึง ดัชนีที่แสดงถึงความสามารถของข้อสอบที่จะจำแนกกลุ่มผู้สอบ ระหว่างผู้ที่มีระดับความสามารถต่ำกว่าระดับความยาก และผู้ที่อยู่สูงกว่าความยาก โดยถือค่าโอกาสที่จะตอบถูกเป็นเกณฑ์ ในทางปฏิบัติจะมีค่า 0 ถึง $+2$ ถ้า $a=0$ แสดงว่าข้อสอบไม่มีอำนาจจำแนก หรืออำนาจจำแนกค่า ถ้า $a = +2$ แสดงว่า ข้อสอบมีค่าอำนาจจำแนกสูง ถ้า a เป็นลบ แสดงว่าข้อสอบไม่ดี จำแนกไม่ได้

11. ค่าการเดา (Item Guessing: c) หมายถึง ดัชนีของข้อสอบที่แสดงถึงความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมากจะมีโอกาสตอบข้อสอบข้อนั้นถูก มีค่าที่เป็นไปได้ อยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกได้ด้วยการเดาแล้ว $C_i = 0$

12. ค่าความสามารถ (Ability: θ) หมายถึง คะแนนจริงของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

13. เกณฑ์ที่เหมาะสมของค่าพารามิเตอร์ของข้อสอบ หมายถึง เกณฑ์ของค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) ของข้อสอบ โดยที่ $a > 0.8$
 $-2.5 \leq b \leq 2.5$ และ $c < 0.3$ (Warm 1978: 52-53)

14. ผู้เชี่ยวชาญแบ่งเป็น - ผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา ได้แก่ อาจารย์ผู้สอนในวิชาที่มีเนื้อหาเกี่ยวกับการวัดและประเมินผลในสถาบันอุดมศึกษาไม่ต่ำกว่า 5 ปี หรือ เป็นผู้สำเร็จการศึกษาในสาขาวิชาการวัดและประเมินผลทางการศึกษาไม่ต่ำกว่าชั้นปริญญาโท

- ผู้เชี่ยวชาญด้านการสอนคณิตศาสตร์ ได้แก่ อาจารย์
ที่ทำหน้าที่สอนวิชาคณิตศาสตร์ในโรงเรียนมัธยม ที่เคยมีประสบการณ์การสอนในระดับมัธยมศึกษา
ปีที่ 3 อย่างน้อย 3 ปี และสำเร็จการศึกษาไม่ต่ำกว่าระดับปริญญาตรี ในสาขาวิชาคณิตศาสตร์

15. ความตรงตามแนวของแบบสอบอิงโดเมน หมายถึง คุณภาพของแบบสอบด้าน
ความตรงเชิงเนื้อหา (Content Validity) และความตรงเชิงโครงสร้าง (Construct
Validity) ความตรงร่วมสมัย (Concurrent Validity) โดยที่

- ความตรงเชิงเนื้อหา หมายถึง คุณสมบัตินี้ของแบบสอบที่สามารถวัดพฤติกรรม
ในขอบเขตเนื้อหาหรือโดเมนที่มุ่งวัดได้ถูกต้อง ตรวจสอบโดยการพิจารณาตัดสินของผู้เชี่ยวชาญ
การสอน

- ความตรงเชิงโครงสร้าง หมายถึง คุณสมบัตินี้ของแบบสอบที่สามารถวัดได้ตรง
ตามโครงสร้างของพฤติกรรมในขอบเขตเนื้อหาที่มุ่งวัดหรือโดเมนตรวจสอบโดยใช้เทคนิคการ
วิเคราะห์ตัวประกอบ

- ความตรงร่วมสมัย หมายถึง คุณสมบัตินี้ของแบบสอบที่แสดงถึงความสัมพันธ์
ระหว่างคะแนนจากการสอบแบบสอบอิงโดเมน กับคะแนนเกณฑ์ ซึ่งได้แก่ เกณฑ์วิชาคณิตศาสตร์
(ค 311) ในภาคเรียนที่ 1 ปีการศึกษา 2531 ของนักเรียนผู้สอบ ตรวจสอบโดยการคำนวณค่า
สัมประสิทธิ์สหสัมพันธ์เพียร์สัน

16. ความเที่ยงตามแนวอิงโดเมน หมายถึง ดัชนีที่แสดงถึงความคงที่ของคะแนน
การสอบของผู้สอบแต่ละบุคคลว่าแตกต่างจากจุดตัดที่ใช้เป็นเกณฑ์อย่างน้อยเพียงใด ซึ่งในการวิจัย
ครั้งนี้ คำนวณจากค่าดัชนีความเชื่อถือได้ (Index of Dependability) ตามวิธีของเบรนนัน
และเคน (Brennan and Kane 1977: 260)

17. ความเที่ยงตามแนวของทฤษฎี IRT หมายถึง ดัชนีที่แสดงถึงความถูกต้อง
แม่นยำในการประมาณค่าความสามารถจริงด้วยผลการตอบแบบสอบ ซึ่งพิจารณาจากค่าฟังก์ชัน
สารสนเทศแบบสอบ (Test Information Function)

18. พฤติกรรมด้านพุทธิพิสัยในวิชาคณิตศาสตร์ หมายถึง จุดประสงค์ทางการเรียน
การสอนที่เน้นพฤติกรรมในด้านสติปัญญา ความรู้ ความคิด และความสามารถในการแก้ปัญหา
ต่าง ๆ ในวิชาคณิตศาสตร์ โดยเจมส์ ดับบลิว วิลสัน (James W. Wilson 1971: 660-664)
ได้แบ่งออกเป็น 4 ชั้น คือ พฤติกรรมด้านความรู้ความจำและการคำนวณอย่างง่าย
(Computation) ความเข้าใจ (Comprehension) การนำไปใช้ (Application) และ
การวิเคราะห์ (Analysis)

19. เอกนาม (Monomial) หมายถึง นิพจน์ที่สามารถเขียนในรูปของการคูณของค่าคงที่กับตัวแปรตั้งแต่ 1 ตัวขึ้นไป โดยที่เลขชี้กำลังของตัวแปรแต่ละตัวเป็นศูนย์หรือจำนวนเต็มบวก

20. พหุนามที่ไม่ใช่เอกนาม หมายถึง นิพจน์ที่สามารถเขียนในรูปการบวกของเอกนามตั้งแต่สองเอกนามขึ้นไป

21. พหุนาม (Polynomial) หมายถึง นิพจน์ที่สามารถเขียนในรูปของการคูณของค่าคงที่กับตัวแปรตั้งแต่ 1 ตัวขึ้นไป โดยที่เลขชี้กำลังของตัวแปรแต่ละตัวเป็นศูนย์หรือจำนวนเต็มบวก และนิพจน์ที่สามารถเขียนในรูปการบวกของเอกนามตั้งแต่สองเอกนามขึ้นไป

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้ตัวอย่างแบบสอบอิงโคเมนวิชาคณิตศาสตร์เรื่องพหุนาม ที่มีมาตรฐานสำหรับใช้ทดสอบนักเรียนชั้นมัธยมศึกษาปีที่ 3 โดยทั่วไปได้ และได้คู่มือการใช้แบบสอบ ซึ่งระบุถึงจุดมุ่งหมายของแบบสอบ คะแนน เกณฑ์ การประมาณค่าความสามารถและคะแนนโคเมนของผู้สอบ การแปลผลคะแนน เป็นต้น ตลอดจนได้แบบฟอร์มการรายงานผลการสอบ ซึ่งผลจากการสอบวัดจะสามารถใช้ตรวจสอบได้ว่า ผู้สอบมีระดับความรู้ความสามารถเป็น เท่าใด เมื่อเทียบกับมวลพฤติกรรมความรู้ทั้งหมดหรือโคเมนที่กำหนดไว้ และความรู้ความสามารถที่มีอยู่ผ่านเกณฑ์การรอบรู้หรือไม่ ตลอดจนเป็นการตรวจสอบว่าผู้เรียนยังบกพร่องในการเรียน เนื้อหาหรือโคเมนใด อันเป็นสารสนเทศที่จำเป็นต่อการปรับปรุงการเรียนการสอนต่อไป

2. เป็นแนวทางสำหรับผู้สนใจศึกษาค้นคว้าเกี่ยวกับการประยุกต์ใช้โมเดลโลจิสติกในการสร้างและพัฒนาแบบสอบต่าง ๆ ให้มีประสิทธิภาพมากที่สุด และเป็นแนวความคิดในการพัฒนาการวัดผลทางการศึกษาในประเทศไทยให้ก้าวหน้าและแพร่หลายต่อไปโดยเฉพาะอย่างยิ่งในระดับกลุ่มโรงเรียนหรือเขตการศึกษา