

# การวัดผลแบบอิงกลุ่มและอิงเกณฑ์

เบญจวรรณ ผ่องแผ้ว

## บทนำ

การสอนและการทดสอบเป็นหน้าที่สำคัญของครูทุกคน การทดสอบเพื่อประเมินว่านักเรียนเรียนรู้มากน้อยเพียงใดหลังจากครูได้สอนบทเรียนต่างๆ ไปแล้ว เป็นหลักปฏิบัติที่ใช้กันทั่วไป ตั้งแต่ระดับประถมจนถึงอุดมศึกษา กระบวนการวัดผลและประเมินผลการเรียนรู้นั้นทำได้หลายวิธี แต่ละวิธีอาจมีปรัชญาเบื้องหลัง และวัตถุประสงค์แตกต่างกันทำให้วิธีการวัดผลและการแปลผลคะแนนแตกต่างกันไปด้วย วิธีการวัดผลที่ใช้กันแพร่หลายมาแต่ดั้งเดิมคือการวัดผลแบบอิงกลุ่ม (Norm-referenced test) คะแนนที่ได้จากการวัดผลวิธีนี้จะนำมาจำแนกนักเรียนเป็นระดับต่างๆ ตามคะแนนที่ได้รับ และจะตัดสินความสำเร็จของการเรียนรู้ โดยนำคะแนนที่ได้มาเปรียบเทียบกับคะแนนของนักเรียนอื่นๆ ในกลุ่ม

ในช่วง 2-3 ทศวรรษที่ผ่านมาได้เกิดการเปลี่ยนแปลงด้านกระบวนการเรียนการสอนครั้งใหญ่ ทำให้เห็นว่าการวัดผลแบบอิงกลุ่มนั้นบอกให้เราทราบว่า นักเรียนสอบได้เป็นอันดับที่เท่าใดในกลุ่ม แต่ไม่สามารถบอกถึงจุดอ่อนหรือความสามารถของนักเรียนในการทำข้อสอบนั้นได้ ไม่เหมาะที่จะนำการทดสอบชนิดนี้มาวัดการเรียนรู้ของแต่ละบุคคล จึงมีการพัฒนาการทดสอบอีกแบบหนึ่งขึ้น คือการวัดผลแบบอิงเกณฑ์ (Criterion-referenced test) การวัดผลแบบอิงเกณฑ์จะนำความสามารถในการเรียนรู้ของนักเรียนมาเปรียบเทียบกับเกณฑ์ที่วางไว้แทนที่จะเปรียบเทียบกับคนอื่น ๆ ในกลุ่มเหมือนการทดสอบแบบแรก

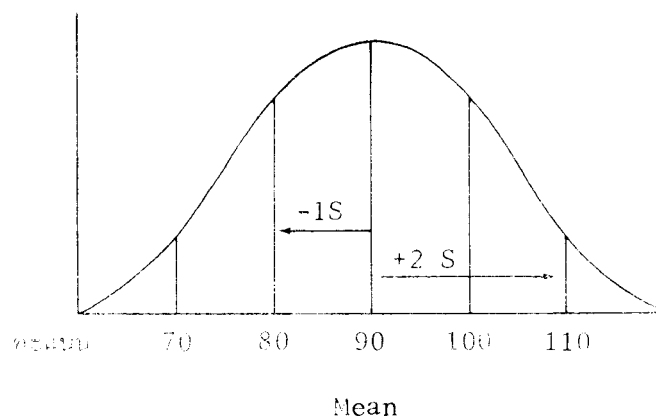
## ความเป็นมาและลักษณะการวัดผลแบบอิงกลุ่ม

การวัดผลแบบอิงกลุ่มและอิงเกณฑ์เป็นการสอบที่มีวิธีการแปลผลคะแนนจากการทดสอบแตกต่างกัน ทั้งนี้เพราะการวัดผลทั้งสองแบบมีพื้นฐานทางปรัชญาที่แตกต่างกัน ปรัชญาการศึกษาแบบดั้งเดิมของเราเชื่อว่าคนเรามีความสามารถต่างกัน ถ้ารวมอบหมายงานชิ้นหนึ่งให้นักเรียนในชั้นไปศึกษา แล้วทดสอบในเวลาต่อมา ผลการสอบจะปรากฏว่า นักเรียนบางคนจะได้คะแนนสูง บางคนได้คะแนนต่ำ แต่นักเรียนส่วนใหญ่จะได้คะแนนปานกลาง นอกจากนี้นักศึกษารุ่นเก่าเชื่อว่า นักเรียนมีสติปัญญาไม่เท่าเทียมกันและบางคนจะไม่ได้ประโยชน์จากการศึกษา ผู้ที่เรียนไม่ได้ก็ไม่จำเป็นต้องเรียน ปรัชญาทางการศึกษานี้เชื่อมั่นในทฤษฎีพันธุกรรมและลักษณะความถ้อย ความเด่นถ้านสติปัญญา จึงเป็นพื้นฐานของทฤษฎีการวัดผลแบบอิงกลุ่ม (Chuhan : 1983)

## ลักษณะของการวัดผลแบบอิงกลุ่ม

ครูทั่วไปรู้จักการวัดผลวิธีนี้เป็นอย่างดี ข้อสอบสัมฤทธิ์ผลส่วนใหญ่เป็นการวัดผลแบบอิงกลุ่ม ผลการสอบจะเป็นเครื่องบอกความสามารถรวม และจะไม่แยกย่อยเป็นพฤติกรรมแต่ละประเภท เนื่องจากการสอบประเภทนี้เป็นไปเพื่อจัดลำดับนักเรียนตามสัมฤทธิ์ผลจากทางการเรียนจากสูงไปหาต่ำ ดังนั้นลักษณะสำคัญของข้อสอบจะประกอบไปด้วยข้อสอบรายข้อที่มีระดับความยากประมาณ 50% มาประกอบเป็นข้อทดสอบ ข้อสอบข้อใดที่นักเรียนตอบถูกต้อง หรือผิดหมดทุกคนจะถูกตัดทิ้ง ทั้งนี้เพราะข้อสอบที่ยากเกินไปหรือง่ายเกินไป ไม่สามารถจำแนกความสามารถที่ต่างกันของนักเรียนได้ (Gronlund : 1976)

เมื่อจัดทำข้อสอบและทำการทดสอบแล้ว จะแปลผลคะแนนโดยการเปรียบเทียบคะแนนนี้กับ norm คำว่า norm คือค่าเฉลี่ย ค่า mean mode และ median ถือว่าเป็น norms ทั้งสิ้น ตัวอย่างคะแนนที่ได้จากการวัดผลแบบอิงกลุ่ม เช่น เปอร์เซนไทล์ คะแนน I.Q. หรือคะแนนมาตรฐาน เป็นต้น เนื่องจากนักเรียนมีความสามารถไม่เท่ากันจึงมีสมมุติฐานเบื้องต้นว่า การกระจายของคะแนนจะออกมาในรูปของโค้งปกติหรือ normal curve นี้ และมีคะแนนสูงและคะแนนต่ำเป็นจำนวนน้อยอยู่หัวท้าย มีสูตรทางเลขคณิตมากมายเพื่ออธิบายและแปลผลคะแนนที่ได้จากการทดสอบ เช่น ช่องคะแนนที่ห่างจาก mean วัดเป็น standard deviation คะแนน -1 standard score หมายความว่าคะแนนที่ได้ต่ำกว่า mean อยู่ 1 standard deviation คะแนน + 2 standard score หมายความว่า เป็นคะแนนที่สูงกว่า mean อยู่ 2 standard deviation (ภาพ 1) (Gilman : 1974)



สมมติว่านักเรียนคนหนึ่งทำข้อสอบชุดหนึ่งถูก 45 ข้อ จากจำนวน 60 ข้อ ถ้าอยากทราบว่าคะแนนนี้มีความหมายอย่างไร ก็ไปเปิดตารางค่าคะแนน 45 จัดเป็นเปอร์เซนไทล์เท่าไร สมมติว่าได้เท่ากับเปอร์เซนไทล์ที่ 86 แสดงว่ามีนักเรียน 14% ได้คะแนนสูงกว่าและ 84% ได้คะแนนต่ำกว่านักเรียนผู้นี้ หรือกล่าวอีกนัยหนึ่ง การทำข้อสอบถูก 45 ข้อจาก 60 ข้อ แสดงว่านักเรียนเรียนรู้ 75% แต่ถ้าจะถามว่าคะแนน 75% หรือเปอร์เซนไทล์ที่ 86 แสดงว่านักเรียนผู้นั้นทำอะไรได้ และทำอะไรไม่ได้บ้าง การวัดผลแบบนี้ก็จะบอกเราไม่ได้มากนัก (Popham : 1980)

## ข้อดีข้อเสียของการวัดผลแบบอิงกลุ่ม

ข้อดีของการวัดผลแบบนี้คือ มีประโยชน์ในการสอบคัดเลือก หรือจำแนกนักเรียนเพื่อจัดกลุ่มการสอน ในกรณีที่เป็นการสอบสัมฤทธิ์ผลสามารถใช้รายงานผลความก้าวหน้าของนักเรียนให้ผู้ปกครองทราบ เป็นเครื่องกำหนดการเลื่อนชั้น หรือใช้เป็นข้อมูลพื้นฐานสำหรับประกอบอาชีพหรือศึกษาต่อไป

ข้อเสียของการวัดผลแบบนี้คือ ทำให้นักเรียนมีความรู้สึกแข่งขันกันสูงทำให้มีความกดดันในด้านการเรียนและการสอบมาก อีกประการหนึ่งคะแนนที่ได้จากการทดสอบนี้ไม่สามารถแจ่มแจ้งให้เราทราบว่านักเรียนสามารถทำอะไรได้บ้าง คะแนนที่เป็นเปอร์เซ็นต์หรือเปอร์เซ็นต์ไทล์ก็มีปัญหาในด้านความเที่ยง นักเรียนที่ได้คะแนน 89 และ 90 อาจมีความแตกต่างกันแง่ตัวเลขและการจัดลำดับ แต่ในแง่ความเที่ยงตรงของการวัดแล้วก็ไม่สามารถแยกได้ว่า นักเรียนทั้งสองมีความสามารถที่แตกต่างกันจริงหรือไม่ และแตกต่างกันอย่างไร

## ความเป็นมาและลักษณะการวัดผลแบบอิงเกณฑ์

การวัดผลแบบอิงเกณฑ์ มีพื้นฐานมาจากทฤษฎีด้านพฤติกรรมของ B.F. Skinner แห่งมหาวิทยาลัยฮาร์วาร์ด ทฤษฎีนี้เชื่อว่าการศึกษาคือสิ่งสากลสำหรับนักเรียนทุกคน ทุกคนสามารถเรียนได้และหน้าที่ของครูก็คือช่วยให้นักเรียนจำนวนมากเล่าเรียน ทฤษฎีนี้เชื่อว่า นักเรียนทุกคนสามารถเรียนบทเรียนใดๆ ก็ได้ หากให้เวลาและโอกาสแก่เขาเพียงพอ ดังนั้น การวัดผลแบบอิงเกณฑ์จึงสนใจความสามารถของนักเรียนแต่ละคนมากกว่าสนใจเปรียบเทียบคะแนนของนักเรียนผู้นั้นกับคนอื่น ๆ

ความเคลื่อนไหวที่ก่อให้เกิดการวัดผลแนวใหม่ขึ้นมีสาเหตุหลายประการดังนี้

1. ความเคลื่อนไหวในการจัดระบบการศึกษาแบบใหม่ เช่น การสอนแบบเอกัตบุคคล (individualized instruction) การสอนแบบโปรแกรม (programmed instruction) การใช้คอมพิวเตอร์ในการเรียนการสอน (computer assisted instruction) ระบบการศึกษาเปิด (open school and university) เป็นเหตุให้การวัดผลแบบอิงกลุ่มไม่เหมาะสมจึงต้องมีการพัฒนาการทดสอบแบบใหม่ขึ้น

2. ระบบการเลือกชั้นเรียนโดยอัตโนมัติตามระบบการศึกษาแบบ Comprehensive School ก่อให้เกิดความวิตกว่านักเรียนได้เลื่อนชั้นโดยไม่มีความรู้พอที่จะเรียนในชั้นใหม่ ถ้าโรงเรียนจะดำเนินนโยบายที่เข้มงวดขึ้นโดยให้นักเรียนสอบผ่านเกณฑ์ที่กำหนด ก็จะแก้ปัญหาได้

3. งานวิจัยเกี่ยวกับการประเมินผลโครงการด้านการศึกษาว่าเป็นไปตามจุดประสงค์ที่วางไว้หรือไม่ ช่วยให้เกิดการวัดผลแบบใหม่ขึ้น

## คำจำกัดความของการวัดผลแบบอิงเกณฑ์

Robert Glaser แห่งมหาวิทยาลัย Pittsburgh เป็นคนแรกที่ชี้ให้เห็นข้อบกพร่องของการใช้ข้อทดสอบอิงกลุ่มกับการเรียนแบบโปรแกรม และได้แสดงให้เห็นความแตกต่างระหว่างการสอบแบบอิงกลุ่ม ซึ่งนำคะแนนของนักเรียนเปรียบเทียบกับของคนอื่น ๆ กับการสอบแบบอิงเกณฑ์ ซึ่งจะเปรียบเทียบนักเรียนกับเกณฑ์มาตรฐานที่

วางไว้ Glaser ได้ให้คำจำกัดความของการวัดผลแบบอิงเกณฑ์ว่า “เป็นการทดสอบที่สร้างขึ้นเพื่อให้สามารถแปลผลการสอบได้ตามเกณฑ์ที่กำหนดไว้” (Glaser : 1963)

หลังจาก Glaser ก็มีผู้ให้คำจำกัดความของการวัดผลแบบอิงเกณฑ์ไว้อีกมากมาย แต่ผู้ที่ให้คำจำกัดความที่กะทัดรัดและมีความหมายครบถ้วนคือ Popham (1978) ซึ่งกล่าวว่าเป็นการทดสอบซึ่งใช้บ่งบอกถึงความสามารถของนักเรียนแต่ละคนตามปริเฉทของพฤติกรรมที่กำหนดไว้ (A criterion-referenced test is used to ascertain an individual's status with respect to a well-defined behaviour domain)

คำว่า ‘a well-defined behaviour domain’ ของ Popham เป็นสิ่งที่ต้องอธิบายขยายความต่อไปว่าหมายถึงอะไร นักทดสอบอาจตีความหมายได้เป็น 2 ประการ คือ พฤติกรรมอันแสดงออกถึงความสามารถในทักษะใดทักษะหนึ่ง เช่น นักเรียนสามารถวิ่ง 100 เมตรได้ เป็นการแสดงว่าปริเฉทนั้นประกอบด้วยพฤติกรรมเดียว แต่ นักทดสอบอีกกลุ่มหนึ่งมีความเห็นว่า a well-defined behaviour domain ประกอบด้วยพฤติกรรมย่อยหลายพฤติกรรมรวมกัน Baker (1974) เป็นผู้ที่สนับสนุนความคิดอย่างหลังนี้ และให้คำจำกัดความของ domain หรือปริเฉทไว้ว่า ปริเฉทประกอบด้วยหน่วยความรู้ย่อย ทักษะ ความเข้าใจ หรือทัศนคติ ซึ่งเป็นลักษณะของเนื้อหาที่นักเรียนต้องเรียนรู้พร้อมทั้งพฤติกรรมที่แสดงว่าได้รับความรู้ตามเป้าหมาย ซึ่งถูกกำหนดและเขียนบรรยายไว้โดยละเอียดรวมอยู่ในปริเฉท กล่าวอีกนัยหนึ่งปริเฉทคือ วัตถุประสงค์ซึ่งระบุบ่งถึงเนื้อหาและพฤติกรรมการเรียนรู้โดยละเอียด

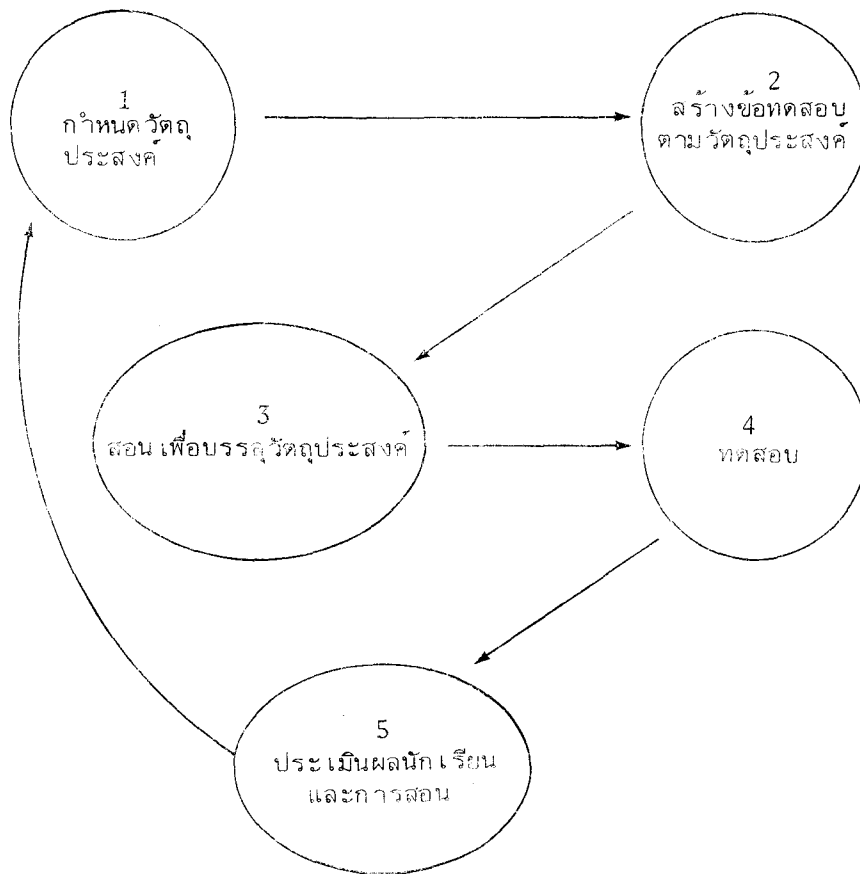
จากคำจำกัดความต่างๆ ที่เกิดขึ้น ทำให้มีการสร้างแบบทดสอบอิงเกณฑ์ ซึ่งมีชื่อเรียกต่างๆ กัน Ebel (1971) ให้ชื่อว่า Content-standard test คะแนนที่ได้จากการสอบจะบอกได้ว่า นักเรียนตอบคำถามจาก task ประเภทใดได้กี่เปอร์เซ็นต์

Osborn (1968) เรียกการทดสอบแบบนี้ว่า universe-defined test เป็นข้อสอบที่สร้างขึ้นจากเนื้อหาสากล (Content universe) ที่กำหนดไว้และคะแนนจากข้อทดสอบจะไม่มีค่าเฉลี่ย

ส่วน Hively (1974) พอใจที่จะเรียกการทดสอบแบบนี้ว่า domain-referenced test หรือการทดสอบแบบอิงปริเฉท

## ขั้นตอนในการสร้างข้อสอบแบบอิงเกณฑ์

การทดสอบแบบอิงเกณฑ์ มีขั้นตอนสำคัญประการแรกคือ กำหนดจุดประสงค์ของการสอนและการทดสอบขึ้น นักทดสอบบางกลุ่มจะเขียนเป็นจุดประสงค์เชิงพฤติกรรม บางกลุ่มก็จะถือว่าเป็นการกำหนดปริเฉทหรือ Domain specification ต่อจากนั้นก็เป็นการสร้างสถานการณ์เพื่อให้พฤติกรรมที่ต้องการเกิดขึ้น ตัวอย่างของสถานการณ์จะถูกเลือกมาเป็นข้อทดสอบ ขั้นตอนต่อไปคือสอนเพื่อให้บรรลุวัตถุประสงค์ เมื่อการสอนสิ้นสุดลง ก็จะทำกรทดสอบเพื่อดูว่า นักเรียนได้คะแนนตามเกณฑ์มาตรฐานที่กำหนดไว้ และการเรียนการสอนได้ตามระดับที่วางไว้หรือไม่ ดังนั้นการทดสอบแบบอิงเกณฑ์จึงไม่ได้วัดผลการเรียนของนักเรียนเป็น summative evaluation แต่อย่างใดแต่ได้ทำหน้าที่ประเมินผลการสอน และสามารถนำข้อมูลนั้นมาปรับปรุงการสอน หรือเป็น formative evaluation ได้ด้วย ขั้นตอนของการวัดผลแบบอิงเกณฑ์ อาจสรุปได้ดังนี้ (C Gilman : 1974)



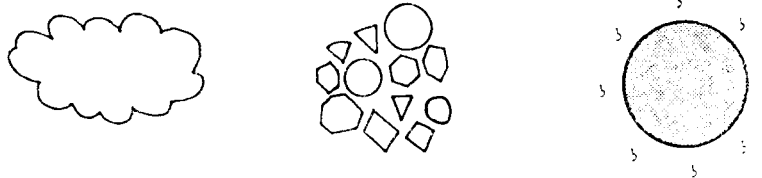
**Mastery test**

จาก Model ของการวัดผลทำให้เกิดการพัฒนาข้อสอบขั้นสูงหนึ่ง ซึ่งเป็นที่รู้จักและใช้กันอย่างแพร่หลายคือ Mastery Test Mastery Test เป็นการทดสอบซึ่งพัฒนาขึ้นสำหรับวัดผลการสอนแบบเอกัตบุคคลหรือ individualized teaching program จุดประสงค์ของการสอบก็เพื่อจำแนกว่านักเรียนได้บรรลุจุดประสงค์ที่กำหนดไว้หรือไม่ เนื่องจากแบบสอนนี้ใช้ทดสอบทักษะเบื้องต้น ปริมาณของพฤติกรรมจึงไม่ซับซ้อน เป็นการวัดผลว่านักเรียนสามารถประกอบทักษะได้อย่างใดอย่างหนึ่งได้หรือไม่ ระดับความยากของการทดสอบนี้จะค่อย ๆ เพิ่มขึ้นตามลำดับ การสอบ Mastery test นี้ได้แก่ข้อสอบวิชาเลขคณิตและการอ่านในระดับประถมศึกษา ได้มีผลงานวิจัยที่สนับสนุนว่าการเรียนและการสอนแบบนี้เป็นผลดีต่อพัฒนาการด้านอารมณ์ และสติปัญญาของนักเรียน นักเรียนมีทัศนคติต่อวิชาเรียนดีขึ้น มีความอยากเรียนมากขึ้นและมีความภาคภูมิใจในตนเอง นักเรียนที่เรียนช้าก็มีความมั่นใจมากขึ้น ปัจจัยที่สำคัญอีกประการหนึ่งของการสอนและการสอบแบบนี้คือ ผู้ที่ถูกกำหนดเมื่อเกิดการล้มเหลวทางการเรียนคือครู ไม่ใช่ นักเรียน

อย่างไรก็ดี Mastery Test ก็มีข้อจำกัดคือจะใช้ได้ผลกับทักษะเบื้องต้น ไม่ซับซ้อน หรือพฤติกรรมเดี่ยวที่ชัดเจน ในกรณีที่เนื้อหาที่มีขอบข่ายกว้าง หรือทักษะที่ซับซ้อนขึ้น ก็ยากที่ Mastery Test จะวัดผลได้

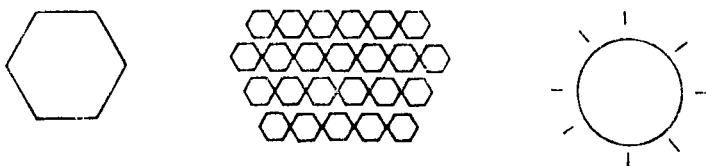
### Domain-Referenced Test

การวัดผลแบบอิงเกณฑ์ตามคำจำกัดความของ Popham คือการวัดผลตามปริเฉทของเนื้อหา หรือพฤติกรรมที่กำหนดให้ ดังนั้นการกำหนดขอบข่ายของปริเฉทของเนื้อหาจึงเป็นสิ่งสำคัญยิ่งสำหรับทักษะในระดับสูงขึ้นไป ทั้งนี้เพราะการแบ่งกลุ่มเนื้อหา และการเขียนจุดประสงค์เชิงพฤติกรรมแบบเดิมมีข้อบกพร่องไม่รัดกุมพอ ทำให้ผู้สอนหรือผู้เขียนข้อทดสอบมีความเข้าใจไม่ตรงกัน เช่นจุดประสงค์เชิงพฤติกรรมกว้าง ๆ ว่า “นักเรียนจะสามารถตั้งชื่อเรื่องจากย่อหน้าที่อ่าน” นั้นจะต้องมีข้อจำกัดอื่น ๆ มาประกอบด้วยว่าเป็นย่อหน้าประเภทไหน ความยาวเท่าไร ฯลฯ เป็นต้น ความไม่รัดกุมของจุดประสงค์เชิงพฤติกรรมอาจทำให้การสอน หรือการสอบไม่สอดคล้องกับวัตถุประสงค์และทำให้การแปลผลการสอบผิดพลาด (Popham : 1980)



กำหนดจุดประสงค์ไม่รัดกุม + ข้อสอบไม่สอดคล้องกับจุดประสงค์ = แปลผลการสอบผิดพลาด

จากข้อบกพร่องดังกล่าว จึงเกิดความพยายามกำหนดวัตถุประสงค์ในรูปของปริเฉทที่ชัดเจนขึ้น เพื่อให้ นักทดสอบสามารถออกข้อสอบรายชื่อที่มีลักษณะเหมือนกันทุกประการได้จาก domain specification อันเดียวกัน เพื่อให้การแปลผลการสอบเป็นไปอย่างถูกต้อง



กำหนดจุดประสงค์ชัดเจน + ข้อสอบสอดคล้องกับจุดประสงค์ = แปลผลการสอบถูกต้อง

จากความล้มเหลวของวัตถุประสงค์เชิงพฤติกรรม ทำให้เกิดความพยายามที่จะเขียนวัตถุประสงค์ที่มีรายละเอียดชัดเจนยิ่งขึ้นเรียกว่า amplified objectives นอกเหนือจากจุดประสงค์และตัวอย่างข้อสอบแล้วก็จะมีการขยายความจุดประสงค์ โดยจะระบุถึงสถานการณ์ในการสอบคัดเลือกและเกณฑ์ความถูกต้องให้ด้วย (ดูตัวอย่างจาก Appendix A) ขอบข่ายของเนื้อหาและพฤติกรรมชัดเจนกว่าการกำหนดจุดประสงค์เชิงพฤติกรรมทั่วไป แต่ยังไม่สามารถแก้ไขปัญหาได้ทั้งหมด ปัญหาเรื่องความกระจัดกระจายความกะทัดรัดในการเขียนข้อกำหนดยังมีอยู่เหมือนเดิม แม้จะน้อยลง

Popham และคณะ ยังคงพยายามต่อไปที่จะกำหนดให้ชัดเจนขึ้น ในที่สุดก็ตกลงกันว่าจะเน้นการวัดพฤติกรรมจำนวนน้อย แต่จะมีการเขียนบรรยายข้อทดสอบนั้นโดยละเอียด วิธีนี้เรียกว่า a limited-focus strategy ลักษณะคล้ายกับ amplified objectives คือจะมีส่วนประกอบต่าง ๆ ดังนี้

**General description** บอกจุดประสงค์ว่าข้อทดสอบวัดอะไร ส่วนนี้บางคนเรียกว่าเป็นจุดประสงค์ของข้อสอบ

**Sample item** คือตัวอย่างข้อสอบให้ไว้เพื่อจุดประสงค์สองประการคือสำหรับผู้ต้องการใช้ข้อทดสอบซึ่งอาจอยากเห็นข้อทดสอบด้วย นอกเหนือจากตารางกำหนดเนื้อหาข้อสอบ และสำหรับผู้ซึ่งต้องการออกข้อสอบที่มีลักษณะเดียวกัน

**Stimulus attributes** คือส่วนซึ่งเป็นตัวกระตุ้นให้เกิดคำตอบ จะมีคำอธิบายว่า stimulus นั้นมีขอบข่ายแค่ไหน และหากจะทำข้อทดสอบแบบเดียวกันนี้อาจทำได้อย่างไร

**Response attributes** เป็นขอบข่ายของคำตอบนักเรียน ซึ่งคำตอบนี้อาจเป็นตัวเลือกที่มีไว้แล้วในข้อสอบ หรือคำตอบที่นักเรียนต้องเขียนเอง ในกรณีที่เป็นคำตอบประเภทเลือกตอบ จะต้องมีการกำหนดตัวเลือกที่ถูกต้องและตัวเลือกที่ผิดให้ด้วย (ดูตัวอย่างจาก Appendix B)

### ข้อดีและข้อเสียของการทดสอบแบบอิงเกณฑ์

ข้อดีของการทดสอบแบบนี้คือ ลดความกดดันด้านแข่งขันในหมู่นักเรียน นักเรียนที่เรียนช้าก็อาจประสบความสำเร็จได้เช่นเดียวกับนักเรียนที่เรียนเก่ง ทำให้พัฒนาการทางด้านอารมณ์ และสังคมของนักเรียนโดยทั่วไปดีขึ้น

ประการที่สอง ครูและนักเรียนทราบเงื่อนไขของการเรียนการสอนตรงกันว่าจะเรียนในเนื้อหาอะไรบ้าง ขอบข่ายของเนื้อหาเป็นอย่างไร วัตถุประสงค์อย่างไร เกณฑ์การสอบผ่านจะต้องทำได้มากน้อยเพียงใด ถ้าทำได้ตามนั้นก็สอบผ่านได้ นับว่าเป็นการเรียนการสอนและการสอบที่มีข้อตกลงชัดเจนตั้งแต่เริ่มต้น

ข้อดีอีกประการหนึ่งคือ เนื่องจากการทดสอบนี้เป็น formative evaluation สำหรับครูด้วย จึงสามารถนำข้อมูลมาปรับปรุงการสอน บทเรียน และอุปกรณ์การสอนให้สมบูรณ์ยิ่งขึ้น นอกจากนี้ครูยังได้ทราบว่า

นักเรียนทำได้ตามวัตถุประสงค์ข้อใดบ้าง และข้อใดที่ยังทำไม่ได้ตามเกณฑ์ช่วยให้ครูทำการสอนซ่อมเสริมได้อย่างตรงจุด และสามารถรายงานผลให้นักเรียนและผู้ปกครองทราบได้ละเอียดกว่าการวัดผลแบบเดิม

ข้อเสียของการวัดผลแบบอิงเกณฑ์คือ ยากที่จะกำหนดจุดประสงค์เชิงพฤติกรรมและขอบข่ายของเนื้อหาให้ครอบคลุมสิ่งที่ต้องการได้ทุกอย่าง นักการศึกษาและนักทดสอบตระหนักว่ามีพฤติกรรมทางการเรียนรู้หลายอย่างที่ยากที่จะเขียนเป็นวัตถุประสงค์และยากที่จะวัด เช่น ความเข้าใจทัศนคติและพัฒนาการด้านต่าง ๆ จึงอาจเป็นอันตรายยิ่งหากโรงเรียนจะเลิกสนใจสิ่งเหล่านี้และจำกัดตัวเองให้ได้คำตอบตาม stimulus หรือสิ่งเร้าที่กำหนดให้

อย่างไรก็ดี นักทดสอบหลายคนเช่น Popham (1980) ได้ตระหนักถึงสิ่งเหล่านี้เป็นอย่างดี และพยายามพัฒนาวิธีกำหนดขอบข่ายของปริศนาที่ชัดเจน Purves และคณะ (1984) ได้กำหนดปริศนาของการเขียนเรียงความไว้โดยละเอียด และน่าสนใจยิ่ง

การเรียนที่จะมีการสอบตามจุดประสงค์แบบอิงเกณฑ์ครูจะต้องทำหน้าที่ของตนอย่างดีเพราะการที่นักเรียนไม่ประสบผลสำเร็จในการเรียนอาจเป็นเพราะครูสอนไม่ดี จัดกระบวนการเรียนไม่เหมาะสม ผลก็คือเกรงว่าอาจจะก่อให้เกิดการสอนเพื่อสอบและข้อสอบจะกลายเป็นตัวกำหนดกิจกรรมการเรียนการสอนของครูบางคนมากเกินไป

ปัจจุบันโรงเรียน ระดับประถมบางแห่งในกรุงเทพฯ ได้นำระบบการเรียนการสอนแบบ Mastery learning มาใช้ทั้งนี้เพราะการเรียนในระดับประถมเป็นการปูพื้นฐาน นักเรียนควรได้ฝึกทักษะและเรียนรู้ในเรื่องต่างๆ ตามหัวข้อที่กำหนด และควรทำได้ตามเกณฑ์ที่วางไว้ เนื่องจากการเรียนในระดับนี้ยังไม่ซับซ้อนเป็นการเรียนและทดสอบไปที่ละเรื่องถ้าสอบไม่ผ่านในเรื่องใดก็จะมีสอนซ่อมเสริมเพื่อให้สอบผ่านได้ตามเกณฑ์ที่กำหนด

การฝึกทักษะระดับสูงบางอย่างก็จำเป็นต้องมีการสอนและการสอบแบบ Mastery test เช่น การฝึกหัดนักบิน ผู้ที่จะเป็นนักบินได้จะต้องทำได้ตามจุดประสงค์และตามเกณฑ์ที่วางไว้ทุกข้อ

วิชาภาษาอังกฤษพื้นฐานหลักสูตรเดิมของสถาบันภาษาเป็นลักษณะ Individualized instruction และมีการทดสอบแบบอิงเกณฑ์ ก็จะมีการทดสอบย่อยหลังจากเรียนไปแล้วและมีการซ่อมเสริมส่วนที่นักเรียนทำได้ไม่ผ่านเกณฑ์ ดังนั้นแบบนี้

ในทางปฏิบัติปัจจุบันการแปลผลคะแนนที่ได้จากการทดสอบ บางครั้งนักทดสอบจะไม่ใช้เกณฑ์หรือกลุ่มแต่เพียงอย่างเดียวเป็นตัวกำหนดระดับคะแนนความสำเร็จหรือความล้มเหลวของนักเรียน โดยเฉพาะในกรณีที่สอนนักเรียนเป็นจำนวนมาก ทั้งนี้เพราะนักทดสอบได้ตระหนักว่า มีองค์ประกอบที่สำคัญอีกประการหนึ่งที่จะมีผลต่อการสอบ นอกเหนือจากเกณฑ์และกลุ่ม คือตัวข้อทดสอบนั่นเอง เราจะแน่ใจได้อย่างไรว่าข้อทดสอบของเราวัดได้แม่นยำตรงตามจุดประสงค์และขอบข่ายเนื้อหาที่เราวางไว้ทุกประการและมีความยากง่ายเท่า ๆ กัน แบบเรียนที่ใช้ในห้องเรียนอย่างแท้จริง ในกรณีนี้คะแนนของกลุ่มจะบอกได้ถึงความยากง่ายของข้อสอบ และจะเป็นข้อมูลให้ครู หรือผู้บริหารใช้ประกอบในการตัดสินใจในการแปลผลคะแนนครั้งสุดท้ายได้



## สรุปและอภิปราย

การวัดผลแบบอิงกลุ่มและอิงเกณฑ์เป็นลักษณะการทดสอบที่ใช้กันแพร่หลายในวงการศึกษาซึ่งมีทั้งข้อดีและข้อบกพร่อง แบบทดสอบทั้งสองชนิดอาจไม่แตกต่างกันเลยเมื่อดูจากลักษณะภายนอก แต่ความแตกต่างจะอยู่ที่วิธีการสร้างข้อสอบและแปลผลคะแนนที่ได้จากข้อสอบนั้น การวัดผลแบบอิงกลุ่มมีประโยชน์ในการสอบคัดเลือก จัดระดับนักเรียน หรือวัดสัมฤทธิ์ผลทางการเรียน แต่เนื่องจากจุดประสงค์สำคัญของข้อสอบคือวัดความแตกต่างระหว่างบุคคล และจัดอันดับความสามารถตามคะแนน โรงเรียนหลายแห่งจึงหันมาใช้ในการทดสอบแบบอิงเกณฑ์เพื่อลดความกดดันในหมู่ผู้เรียน อย่างไรก็ตาม แม้ว่าการทดสอบแบบอิงเกณฑ์จะแก้ปัญหาคือการเรียนการสอนได้หลายอย่างเช่น การแข่งขันในหมู่ผู้เรียน สามารถบอกขอบข่ายความสามารถนักเรียนได้โดยไม่ต้องเปรียบเทียบกับผู้อื่น และสามารถวัดผลการสอนได้เวลาเดียวกัน แต่การวัดผลแบบนี้ก็ยังไม่สมบูรณ์ ปัญหาใหม่ที่อาจเกิดจากการวัดผลแบบนี้คือการสอนเพื่อสอบ และในแง่การสร้างข้อสอบก็มีเรื่องที่จะต้องกันคว่าเพื่อปรับปรุงอีกหลายด้าน เช่น การกำหนดขอบข่ายของจุดมุ่งหมาย และการเขียนข้อสอบให้ตรงตามจุดมุ่งหมาย เป็นต้น

จึงเป็นสิ่งจำเป็นที่ผู้สอนและผู้บริหารที่จะทราบถึงลักษณะของข้อสอบทั้งสองประเภทเพื่อสามารถเลือกใช้ได้ตามความเหมาะสม และจำเป็นที่จะต้องมีการค้นคว้าวิจัยต่อไปเพื่อให้การสร้างข้อทดสอบใช้ได้ผลสมบูรณ์ยิ่งขึ้น

### References

- Baker, Eva L. (1974). "Beyond Objectives : Domain-Referenced Tests for Evaluation and Educational Improvement." *Domain-Referenced Testing*, Englewood Cliffs, NG : Educational Technology Publications.
- Chuahan, SS. (1983). *Innovations in Teaching-Learning Process*, 2nd ed., Vikas Publishing House.
- Ebel, R.L. (1971). "Content Standard Test Scores." *Educational and Psychological Measurement*, 22, 15-25.
- Gilman, David A. (1974). *Alternatives to Tests, Marks and Class Rank*, Terre Haute, ISU Curriculum Research and Development Center.
- Glaser, R. (1963). "Instructional Technology and the Measurement of Learning Outcomes." *American Psychologist*, 18, 519-21.
- Gronlund, Norman E. (1976). *Measurement and Evaluation in Teaching*, 3rd ed., Collier Mc. Millan International Editions.
- Hively, Wells. (1974). *Domain Reference Testing*. Englewood Cliffs, NGJ. : Educational Technology Publications.
- Osborn, H.G. (1968). "Item Sampling for Achievement Testing." *Educational and Psychological Measurement*. 28, 95-104.
- Popham, W.G. (1978). *Criterion. Referenced Measurement*, Englewood Cliffs, NGJ. : Prentice Hall.
- (1980). "Domain Specification Strategies" in A. Berk (Ed.), *Criterion-Referenced Measurement. The State of Art*, Baltimore.: John Hopkins University Press.
- Purves A.C. et al (1984). "Towards a Domain-Referenced System for Classifying Composition Assignments." *Research in the Teaching of English*, 18, 385-415.

### Appendix A

An Illustrative IOX Amplified Objective for a Third-Grade Level Reading  
Comprehension Skill

#### DETERMINING SEQUENCE FROM TENSE AND WORDS THAT SIGNAL ORDER

##### Objective :

The student will correctly identify the sequence of three sentences by determining order from tense and word words that signify order.

##### Sample Item :

Directions. Read the three sentences. Then mark an "X" next to the answer that arranges the sentences in the proper order.

##### Example :

A. Once there were only candles for lighting the home.

B. Later there were dim electric lights.

C. Tesla thought of a way to make the electric lights brighter,

\_\_\_\_\_a) A, C, B

\_\_\_\_\_b) A, B, C

\_\_\_\_\_c) C, A, B

**Amplified Objective :****Testing Situation.**

1. The student will be given three sentences and will identify their proper sequence on the basis of verb tenses and signal words.
2. Three sentences containing signal words and/or changes in verb tense will be provided.
3. Vocabulary will be familiar to the third grader.

**Response Alternatives.**

1. Three possible orderings of the sentence will be given.
2. At least one distractor should not consist of a random ordering. It should maintain the first event as first, varying only the second and third events.
3. The other distractor may be any other incorrect ordering of the events.

**Criterion of Correctness.**

The correct answer will be the order that can be determined on the basis of one of the following :

1. Words that signify sequence, e.g., afterwards, finally, then, before, during, now, next, lastly, later, earlier, meanwhile, long ago, once.
2. Verb tense (future, past, present)

**Appendix B**

An Illustrative Set of Criterion-Referenced Test Specifications  
for a High School Minimum Competency Test in Reading

**DETERMINING MAIN IDEAS****General Description**

The student will be presented with a factual selection such as a newspaper or magazine article or a passage from a consumer guide or general-interest book. After reading that selection, the student will determine which one of four choices contains the best statement of the main idea of the selection. This statement will be entirely accurate as well as the most comprehensive of the choices given.

**Sample Item**

Directions. Read the selections in the boxes below. Answer the questions about their main ideas.

**THE COLD FACTS**

Had you lived in ancient Rome you might have relieved the symptoms of a common cold by sipping a broth made from soaking an onion in warm water. In Colonial America you might have relied on an herbal concoction made from sage, buckthorn, goldenseal, or bloodroot plants. In Grandma's time, lemon and honey was a favorite cold remedy, or in extreme cases, a hot toddy

laced with rum. Today, if you don't have an old reliable remedy to fall back on, you might take one of thousands of drug preparations available without prescription. Some contain ingredients much like the folk medicines of the past; others are made with complex chemical creations. Old or new, simple or complex, many of these products will relieve some cold symptoms, such as a stopped-up nose or a hacking cough. But not a single one of them will prevent, cure, or even shorten the course of the common cold.

1. Which one of the following is the best statement of the main idea of the article you just read?
  - a. Old-fashioned herbal remedies are more effective than modern medicines.
  - b. There are many kinds of relief, but no real cures, for the common cold.
  - c. Some of today's cold preparations contain ingredients much like those found in folk remedies of the past.
  - d. Americans spend millions of dollars a year on cold remedies.

### **Stimulus Attributes**

1. Each item will consist of a reading selection followed by the question "Which one of the following is the best statement of the main idea of the (article selection) you just read?" Eligible reading selections include adaptations of passages from factual texts such as general-interest books and consumer guides and pamphlets. Care should be taken to pick selections of particular interest to young adults and to avoid selections which may in the near future appear dated. Each reading selection will be titled, will be at least one paragraph long, and will contain from 125-250 words. Not more than 1,000 words of reading material can be tested in any set of five items. At least two of the five items in any set of five items must contain reading selections that are more than one paragraph long.
2. If necessary, the following modifications may be made to a selection used for testing :
  - a. A title may be added if the selection does not have one, or if the selection represents a section of a longer piece whose title would not be applicable to the excerpt. If a title is added, it should be composed of a brief, interest getting and/or summarizing group of words.
  - b. A selection may be shortened, but only if the segment which is to be used for testing makes sense and stands as a complete unit of thought without the parts which have been omitted. If necessary, minor editing can be done to a reading selection which represents a shortening of a longer piece, but this editing should be for the purposes of clarity and continuity only, and not for the purposes of increasing or decreasing the difficulty level, or changing the content of the text.
3. Reading selections used for testing should not exceed a 9th grade reading level, as judged by the Fry readability formula.

### **Response Attributes**

1. A set of four single-sentence response alternatives will follow each reading selection and its accompanying question. All of these statements must plausibly relate to the content of the reading selection, either by reiterating or paraphrasing portions of that selection or by building upon a word or idea contained in the selection.
2. The three incorrect response alternatives will each be based upon a lack of one of the two characteristics needed by a correct main idea statement: accuracy and appropriate scope. A correct main idea statement must be accurate in that everything it states can be verified in the text it describes. It must have appropriate scope in that it encompasses all of the most important points discussed in the text that it describes.

3. A distractor exemplifies a lack of accuracy when it does any one or more of three things :
  - a. Makes a statement contradicted by information in the text.
  - b. Makes a statement unsupported by information in the text. (Such a statement would be capable of verification or contradiction if the appropriate information were available.)
  - c. Makes a statement incapable of verification or contradiction; that is, a statement of opinion. (Such statements include value judgments on the importance or worth of anything mentioned in the text.)
4. A distractor exemplifies a lack of appropriate scope when it does one of two things :
  - a. Makes a statement that is too narrow in its scope. That is, the statement does not account for all of the important details contained in the text.
  - b. Makes a statement that is too broad in its scope. That is, the statement is more general than it needs to be in order to account for all of the important details contained in the text.
5. The important points which must be included in a main idea statement are those details which are emphasized in the text by structural, semantical, and rhetorical means such as placement in a position of emphasis, repetition, synonymous rephrasing, and elaboration. Whether any given main idea statement contains all of the important points that it should is always debatable rather than indisputable. The nature of the question asked on this test, i.e.; select the best main idea statement from among those given, attempts to account for this quality of relative rather than absolute correctness.