



วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้องสำหรับการวิจัยฉบับนี้ จัดแบ่งเนื้อหาออกได้เป็น 2 ประเด็นหลัก คือ

1. พัฒนาการของแบบสอบอิงเกณฑ์
 2. เทคนิคการวิเคราะห์ตามลำดับชั้น
- โดยในแต่ละประเด็นมีสาระสำคัญดังต่อไปนี้

1. พัฒนาการของแบบสอบอิงเกณฑ์

พัฒนาการของแบบสอบอิงเกณฑ์นั้นจากสมัยแนวความคิดของฟลานากานในปี 1950 ถึงปี 1962 เกลเซอร์ในปี 1963 โปแฟมและฮูเสคในปี 1969 การเสนอบทความ การรายงานผลที่ได้จากการค้นคว้าวิจัยเกี่ยวกับแบบสอบอิงเกณฑ์มีเป็นจำนวนมาก พัฒนาการในบางส่วนเกี่ยวกับแบบสอบอิงเกณฑ์ที่ค่อนข้างจะมีความสมบูรณ์หรือมีความเข้าใจตรงกันในหมู่นักวิชาการ พัฒนาการในบางเรื่องก็ยังไม่ถึงจุดสมบูรณ์อันเป็นที่ยอมรับโดยทั่วไปในหมู่นักวิชาการ ทั้งบทความต่างๆก็มีความกระจัดกระจายอยู่มาก การจกัรวบรวมแนวคิดให้เป็นหมวดหมู่เพื่อให้เกิดความสะดวกในการค้นคว้าหรือทำความเข้าใจยังมีอยู่น้อย การเสนอวรรณคดีที่เกี่ยวข้องในประเด็นนี้จึงใคร่ขอแยกประเด็นย่อยต่างๆที่จะกล่าวถึงออกเป็นข้อๆ และจะนำเสนอตามลำดับ ดังต่อไปนี้

- 1.1 แนวคิดทั่วไปและความหมายของแบบสอบอิงเกณฑ์
- 1.2 การพัฒนาวิธีการสร้างและวิเคราะห์ข้อสอบของแบบสอบอิงเกณฑ์
- 1.3 แนวคิดของการกำหนดเกณฑ์มาตรฐาน
- 1.4 ความตรงเชิงประจักษ์ของการกำหนดเกณฑ์มาตรฐาน
- 1.5 การหาค่าความเที่ยงของแบบสอบอิงเกณฑ์
- 1.6 ปัญหาการกำหนดความยาวของแบบสอบอิงเกณฑ์
- 1.7 พัฒนาการของแบบสอบอิงเกณฑ์ในประเทศไทย

1.1 แนวคิดทั่วไปและความหมายของแบบสอบอิงเกณฑ์

ตามประวัติแล้ว คนแรกที่ใช้คำว่า "การวัดในระบบอิงเกณฑ์" (Criterion - Referenced Measurement) ในความหมายของการอิงเกณฑ์คือ เกลเซอร์ ในปี 1963 แต่ก่อนที่จะถึงแนวคิดของเกลเซอร์ ก็มีแนวคิดของฟลานากานและอีเบลมาก่อน โดยอีเบล (Ebel in Mehrens and Ebel 1972 - 62) ได้อ้างถึงแนวคิดของฟลานากานที่ว่า

- คะแนนจากแบบสอบจะมีความหมายและมีคุณค่าก็ต่อเมื่อคะแนนเหล่านั้นสามารถแปลความหมายได้ในเทอมของศักยภาพ ความสามารถ และความสำเร็จในกิจกรรมของการศึกษาที่มีความสำคัญ"

ในปี 1962 อีเบลได้เขียนบทความเรื่อง "คะแนนจากแบบสอบที่กำหนดมาตรฐานของเนื้อหา" (Content Standard Test Score) โดยอีเบลได้กล่าวถึงความหมายของคะแนนไว้ว่า

- ความหมายของคำว่าคะแนนจากแบบสอบที่กำหนดมาตรฐานในเนื้อหาของเนื้อหาหมายถึงจำนวนเปอร์เซ็นต์ของงานที่ได้รับ การสุ่มขึ้นมาจากขอบเขตที่ได้รับ การกำหนดไว้ซึ่งเด็กนักเรียนสามารถทำได้ "

จากแนวคิดของอีเบลข้างต้น จุดเน้นที่สำคัญคือการที่จะต้องนิยามขอบเขตของเนื้อหาหรือโดเมนที่ชัดเจนเป็นประการสำคัญ ต่อมา เกลเซอร์ (Glaser 1963 - 519) ได้เขียนไว้ว่า

- ถ้าให้ความสามารถของการเรียนรู้ในเรื่องใดเรื่องหนึ่งมีลักษณะเป็นความต่อเนื่องจากการที่ผู้สอบไม่มีความรู้เลย ไปจนถึงการที่ผู้สอบมีความรู้ในเรื่องนั้นอย่างสมบูรณ์ และความสามารถของผู้สอบจะได้รับการชี้บ่งว่าอยู่ตรงจุดใดในช่วงความต่อเนื่องนั้น ทั้งนี้แล้ว การตรวจสอบความสามารถของผู้สอบว่าอยู่ตรงระดับใดเมื่อเทียบกับเกณฑ์มาตรฐานที่กำหนดขึ้น ก็จะเรียกได้ว่าเป็นระบบของการวัดผลแบบอิงเกณฑ์ "

จากข้อเขียนของเกลเซอร์ถึงกลาวันซ์ให้เห็นว่า การสอบแบบอิงเกณฑ์เน้นการเปรียบเทียบความสามารถของผู้เรียนกับเกณฑ์มาตรฐานที่กำหนด ซึ่งจะแตกต่างไปจากแนวคิดของการสอบแบบอิงกลุ่มที่มุ่งเน้นการเปรียบเทียบระดับความสามารถของผู้เรียนด้วยกันเองภายในกลุ่ม ซึ่งเกลเซอร์มีความเห็นว่า การกระทำเช่นนั้นก็ทำให้แบบสอบวัดผลสัมฤทธิ์ทางการเรียนไม่มีความแตกต่างไปจากการวัดความถนัด (Glaser 1963 : 519) ไฮวาลี (Hively 1974) ได้กล่าวถึงลักษณะสำคัญของคำว่า "โคเมน" ไว้ว่า จะต้องประกอบไปด้วยข้อสอบที่เป็นตัวแทนของภาระงานทั้งหมด (Universe) ทั้งนี้จากแบบสอบที่สร้างขึ้นถ้าผู้สอบทำข้อสอบและได้รับการพิจารณาว่าเป็นผู้รอบรู้แล้ว จะมีความหมายว่ามีความรู้ในเนื้อหาที่กำหนดทั้งหมด ขอบเขตของเนื้อหาที่กำหนดขึ้นมานั้นอาจจะวัดในวัตถุประสงค์เดียว หรืออาจจะวัดในวัตถุประสงค์หลายๆข้อก็ได้ แต่ประการสำคัญจะอยู่ตรงประเด็นที่ว่า ขอบเขตของเนื้อหาจะต้องได้รับการกำหนดให้ชัดเจน

แบบสอบอิงเกณฑ์จึงได้รับการนิยามให้หมายถึงแบบสอบที่มุ่งวัดเพื่อที่จะได้ทราบถึงระดับสมรรถภาพของบุคคลภายใต้ขอบเขตของเนื้อหาที่ได้รับการกำหนดขึ้น หรือดังที่ไปพัม (Popham 1978 : 93) ได้ให้คำจำกัดความไว้ว่า

" แบบสอบอิงเกณฑ์ คือ แบบสอบที่ใช้เพื่อค้นหาสมรรถภาพของบุคคลภายใต้ขอบเขตของเนื้อหาที่ได้รับการกำหนดอย่างชัดเจน "

การแปลความหมายของคะแนนที่ได้จากแบบสอบอิงเกณฑ์มุ่งเน้นตรงการยึดเอาระดับความสามารถของแต่ละบุคคลกระทำใ้เมื่อเทียบกับเกณฑ์มาตรฐานที่กำหนดขึ้นเป็นหลัก โดยไม่ต้องอาศัยระดับความสามารถของบุคคลอื่นที่ทำแบบสอบชุดเดียวกันมาใช้เป็นมาตรฐานในการแปลความหมายของคะแนนที่ได้ ลินด์วอลล์และนิทโก (Lindwall and Nitko in Glass 1978 : 241 - 242) ได้ระบุถึงลักษณะของแบบสอบอิงเกณฑ์ไว้ 4 ประการ คือ

1. มีการกำหนดพฤติกรรมเฉพาะที่แสดงออกถึงระดับผลสัมฤทธิ์ใ้ได้อย่างเด่นชัดเท่าที่จะเป็นไ้ก่อนที่จะมีการดำเนินการสร้างแบบสอบ

2. พฤติกรรมแต่ละประเภทจะได้รับการระบุในรูปของสถานการณ์ของการสอบ (กล่าวคือจะต้องระบุถึงลักษณะของข้อสอบ ภาระงานในการสอบ) ที่จะทำให้พฤติกรรมทั้งหมดซึ่งต้องการศึกษาสามารถปรากฏออกมาได้

3. เมื่อพฤติกรรมที่ต้องการให้แสดงออกและสถานการณ์ของการสอบได้รับการกำหนดแล้ว ก็จะมีการวางแผนการสอบ เพื่อที่จะได้ภาวะการสอบในรูปแบบใดรูปแบบหนึ่งที่จะเป็นตัวแทนภาระงานทั้งหมดนั้นได้

4. คะแนนที่ได้จากการสอบจะสามารถแปลความหมายได้อย่างเป็นปรนัย และมีความหมายในตัวของมันเอง ในการที่จะแสดงถึงคุณลักษณะของผู้สอบแต่ละคนในประเด็นที่มุ่งวัดทั้งหมด

การอภิปรายและถกเถียงขอบเขตและความหมายที่แท้จริงของแบบสอบอิงเกณฑ์ว่าควรจะเป็นอย่างไรมีอย่างกว้างขวาง จากข้อเขียนของเกลเซอร์ในปี 1963 และไฮวีย์ในปี 1974 เมื่อพิจารณาอย่างลึกซึ้งแล้วก็จะพบความแตกต่างของจุดนั้น กล่าวคือ เกลเซอร์เห็นว่า การพัฒนาและแปลความหมายของแบบสอบอิงเกณฑ์จะเป็นไปเพื่อตรวจสอบว่าผู้เรียนมีความสามารถระดับใดเมื่อเทียบกับ เกณฑ์มาตรฐานที่กำหนดขึ้น ส่วนไฮวีย์เห็นว่าในการพัฒนาแบบสอบประเภทนี้จะต้องมีการกำหนดขอบเขตของเนื้อหาให้เด่นชัดก่อน และการสร้างแบบสอบจะต้องสร้างให้ได้แบบสอบที่ประกอบด้วยข้อสอบที่เป็นตัวแทนของภาระงานทั้งหมด เป็นหลักการสำคัญ

ประเด็นของการโต้แย้งที่สำคัญประการหนึ่งก็คือการเลือกใช้คำเรียกแบบสอบระหว่างคำว่า "อิงเกณฑ์" (Criterion - Referenced) กับ "อิงโดเมน" (Domain - Referenced) โดยเมื่อพิจารณาจากข้อเขียนต่างๆ แล้วพบว่า ในระยะแรกคำที่ปรากฏมาก่อนคือคำว่า "อิงเกณฑ์" นักคิดไม่ว่าจะเป็นอีเบลหรือ เกลเซอร์ก็มีความเห็นตรงกัน กล่าวคือ การแปลความหมายของคะแนนจากแบบวัดผลสัมฤทธิ์ทางการเรียนนั้นควรจะยึดเอาเกณฑ์มาตรฐานหรือร้อยละของข้อคำถามที่ตอบถูกเป็นหลักในการแปลความหมาย ไม่ใช่ยึดเอาคะแนนในกลุ่มผู้สอบด้วยกันเป็นหลักในการแปล โดยทั้งอีเบลและ เกลเซอร์ก็ได้กล่าวไว้เกี่ยวกับการแปลความหมายของคะแนนที่ได้จากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนนั้นจะต้องยึดเอาขอบเขตของเนื้อหาที่กำหนดขึ้น เป็นหลักสำคัญ ระยะต่อมาปรากฏในข้อเขียนของไฮวีย์ (Hively 1974 : 5) นิตโก (Nitko in Traub 1979 : 50) และเบรนนัน (Brennan 1981 : 5) กล่าวไว้ตรงกันคือคนทั่วไปมักจะเข้าใจความหมายของคำว่า "อิงเกณฑ์" ในลักษณะที่ว่า แบบสอบใดๆก็ตาม เมื่อมีการกำหนด เกณฑ์มาตรฐานขึ้นมาใช้กับแบบสอบนั้นๆ ก็จะถือว่าเป็น "แบบสอบอิงเกณฑ์" โดยไม่ได้คำนึงถึง

ขอบเขตของเนื้อหาและวิธีการที่จะสร้างข้อสอบให้เป็นตัวแทนของเนื้อหาทุกอย่างใด เหตุนี้ นักศึกษากลุ่มหนึ่ง เช่น ไฮวี่ลี และ เบรนนัน จึงหันมาใช้คำว่า "อิงโคเมน" แทน ทั้งนี้เพื่อความสำคัญหรือจุดเน้นของการพัฒนาแบบสอบมาอยู่ตรงการที่จะคงมีการกำหนดขอบเขตของเนื้อหาเป็นประการสำคัญ แต่ขณะเดียวกันก็ยังคงมีนักศึกษากลุ่มหนึ่ง เช่น โปแฟม, แอมเบิลตัน, ฮาลาไคนา และ รอยด์ ก็ยังคงใช้คำว่า "อิงเคมท์" เช่นเดิม

นิตโก (Nitko 1980 : 466) ได้อภิปรายถึงประเด็นการเรียกและสื่อความหมายของคำว่า "อิงเคมท์" กับ "อิงโคเมน" ไว้เช่นกัน และเพื่อการสื่อความหมายที่ชัดเจน นิตโกจึงได้เสนอวิธีการจัดแบ่งประเภทของแบบสอบอิงเคมท์โดยยึดเอาลักษณะของการสร้างเป็นสำคัญ โดยนิตโกได้จำแนกแบบสอบอิงเคมท์ออกเป็น 4 ประเภท คือ

1. แบบสอบที่ไม่ได้มีการกำหนดขอบเขตของเนื้อหาไว้ให้ชัดเจน (Undefined Domains)
2. แบบสอบที่มีการกำหนดขอบเขตของเนื้อหาขึ้นไว้อย่างหยابๆ (Ill-defined Domains)
3. แบบสอบที่มีการกำหนดขอบเขตของเนื้อหา แต่ชัดเจนแต่ไม่ได้มีลักษณะของการเรียงลำดับเนื้อหา (Well-defined but unordered Domains) ลักษณะของแบบสอบในประเภทนี้ จะมีการกำหนดขอบเขตของเนื้อหาอย่างชัดเจน แต่ทว่าข้อสอบในแบบสอบยังไม่สามารถหาเกณฑ์มาจัดเรียงลำดับให้มีความหมายเฉพาะได้ ตัวอย่างเช่น แบบสอบที่สร้างขึ้นเพื่อวัดผลสัมฤทธิ์ทางการเรียนในวัตถุประสงค์ไว้ทุกวัตถุประสงค์หนึ่ง นิตโกกล่าวไว้ว่า ความจริงแล้วคำว่า "แบบสอบอิงเคมท์" ที่ใช้กันทั่วไป จะมีความหมายใกล้เคียงกับแบบสอบที่จัดเข้าในประเภทนี้มากที่สุด
4. แบบสอบที่มีการกำหนดขอบเขตของเนื้อหาอย่างชัดเจน และมีการเรียงลำดับ (Well-defined and ordered Domains) ลักษณะของแบบสอบในประเภทนี้นอกจากจะมีการกำหนดขอบเขตของเนื้อหาให้ชัดเจนแล้ว ข้อสอบในแบบสอบยังสามารถนำมาเรียงหรือจัดลำดับให้มีความหมาย โดยใช้เกณฑ์ใดเกณฑ์หนึ่งในการจัด เช่น การเรียงตามลำดับค่าความยาก การเรียงตามความซับซ้อนของทักษะ การเรียงตามค่าประจำข้อสอบที่ได้มาจากการวิเคราะห์ทฤษฎีการตอบสนองข้อสอบ

ชูเมคเกอร์ (Shoemaker 1975 : 129 - 131) ได้สรุปไว้ว่า ข้อสอบประจำโคเมน หมายถึง ข้อสอบในขอบเขตเนื้อหาทั้งหมดที่สามารถสร้างขึ้นได้ ทั้งนี้โปรแกรมการ

สอนจะต้องเป็นสื่อหรือวิธีที่จะทำให้ผู้เรียนสามารถตอบข้อคำถามต่างๆ อยู่ในโคเมนที่กำหนดขึ้นได้ รอยด์ และ ฮาลาดินา (Roid and Haladyna 1982 : 230 - 231) ได้ให้ความเห็นว่า โดยทั่วไปแล้วมีความเป็นไปได้ที่ในเนื้อหาของลักษณะอาจจะกำหนดขอบเขตที่เกินขีดไต่ยาก แต่แนวคิดมีอยู่ว่าถ้าผู้สร้างแบบสอบสามารถกำหนดขอบเขตของเนื้อหาให้เกินขีดไต่เท่าใด ก็จะมีข้อดีในการที่จะกำหนดมาตรการในการวัดผลสัมฤทธิ์ทางการเรียนที่จะเกิดขึ้นได้อย่างชัดแจ้ง ทั้งนี้การกำหนดขอบเขตของเนื้อหาให้ชัดเจนมีประโยชน์ดังต่อไปนี้

1. ก่อให้เกิดความสอดคล้องอย่างชัดแจ้งระหว่างวัตถุประสงค์ วิธีการสอน และ ข้อสอบในแบบสอบ

2. มีวิธีการที่จะสร้างข้อสอบซึ่งสามารถบรรยายให้ทุกคนทราบได้ และสามารถสร้างข้อสอบในรูปแบบเดียวกันออกมาได้เป็นจำนวนมาก หรือสามารถบรรยายลักษณะของแบบสอบได้อย่างชัดแจ้ง

3. ควบคุมให้ผู้เขียนข้อสอบต่างคนกันสามารถเขียนข้อสอบให้วัดได้ความจุประสงค์ที่ต้องการ

4. ก่อให้เกิดความชัดแจ้งในการที่จะแปลความหมายของคะแนนที่ได้จากการสอบเมื่อเทียบกับขอบเขตของความรู้ทั้งหมดที่ต้องการสรุปอ้างอิง ไปถึง ได้อย่างน่าเชื่อถือ

สาระสำคัญโดยสรุปในประเด็นแนวคิดทั่วไปและความหมายของแบบสอบอิงเกณฑ์มีอยู่ว่า จากปี 1962 เป็นต้นมา โคนิกคิกกลุ่มหนึ่งเห็นว่าการพัฒนาแบบสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตลอดจนทั้งการแปลความหมายของคะแนนที่ได้ควรจะมีมาตรการที่แตกต่างไปจากการคำนวณงานตามแนวคิดของการประเมินผลแบบอิงกลุ่มที่มีมาก่อน โดยนักศึกษากลุ่มนี้พยายามชี้ถึงประเด็นของการแปลความหมายของคะแนนที่ได้โดยการกำหนดเกณฑ์มาตรฐาน ตลอดจนทั้งการกำหนดขอบเขตของเนื้อหาที่จะใช้ เป็นหลักในการสร้างแบบสอบ ทั้งนักคิดในกลุ่มนี้ยังแบ่งออกได้เป็นสองกลุ่มย่อยที่มีจุดเน้นในการพัฒนาแบบสอบต่างกัน กล่าวคือ กลุ่มหนึ่งเห็นว่าจุดเน้นของแบบสอบที่สร้างจะอยู่ที่ตรงการกำหนดเกณฑ์มาตรฐานของแบบสอบที่สร้างขึ้นจากขอบเขตของเนื้อหาที่กำหนด และเรียกแบบสอบที่พัฒนาขึ้นมาไว้ว่า "แบบสอบอิงเกณฑ์" ส่วนนักคิดอีกกลุ่มหนึ่งมีความเห็นว่าในการพัฒนาแบบสอบประเมินผลนั้นจะต้องมีการกำหนดขอบเขตของเนื้อหาขึ้นให้ชัดแจ้งเสียก่อน เป็นประเด็นสำคัญ ส่วนเกณฑ์มาตรฐานนั้น เป็นผลที่เกิดขึ้นทีหลัง เมื่อมีการพัฒนาแบบสอบที่มีความเป็นตัวแทนของเนื้อหาขึ้นได้แล้ว และเรียกแบบสอบที่พัฒนาขึ้นมาว่า "แบบสอบอิงโคเมน"

1.2 การพัฒนาวิธีการสร้างและการวิเคราะห์ข้อสอบ ในแบบสอบอิงเกณฑ์

ตามแนวคิดของแบบสอบอิงเกณฑ์ นอกจากจะต้องกำหนดวัตถุประสงค์และขอบเขตของเนื้อหาที่ต้องการจะสรุปพวกพิง ไปถึงให้ชัดเจนแล้ว ยังมีความจำเป็นที่จะต้องระบุถึงขอบเขตประชากรของข้อสอบให้เด่นชัดอีกด้วย แนวคิดพื้นฐานดั้งเดิมง่าย ๆ ที่จะทำได้ข้อสอบเป็นจำนวนมากพอที่จะถือได้ว่าครอบคลุมเนื้อหาทั้งหมดที่ต้องการศึกษา คือ การสร้างธนาคารข้อสอบ (Shoemaker 1975 : 133) โดยอาจจะใช้วิธีระดมนักเขียนข้อสอบมาช่วยกันเขียนและเก็บรวบรวมไว้ อย่างไรก็ตามวิธีการดังกล่าวจะมีปัญหาในประเด็นที่ว่า ผู้สร้างข้อสอบต่างคนกัน แม้จะได้รับการมอบหมายให้สร้างข้อสอบที่วัดในวัตถุประสงค์เดียวกัน แต่ข้อสอบที่เขียนออกมา ก็อาจมีความแตกต่างกันอยู่มาก จึงได้มีการพยายามพัฒนาแนวคิดการสร้างข้อสอบที่มีความเป็นปรนัยมากขึ้น มิลล์แมน (Millman in Popham 1974) ได้เคยสรุปถึงเทคนิคต่างๆ ในการสร้างข้อสอบในแบบสอบอิงเกณฑ์ที่มีความเป็นปรนัยไว้ ต่อมา รอยด์ และ ฮาลาดินา (Roid and Haladyna 1980) ได้สรุปถึงแนวคิดของการสร้างข้อสอบที่เรียกว่าเป็น "เทคโนโลยีของการเขียนข้อสอบ" ไว้ โดยมีเทคนิคใหญ่ๆ พอสรุปได้เป็น 3 แนวทาง คือ

(ก) วิธีการกำหนดวัตถุประสงค์พื้นฐาน (Objective-based Methods)

เทคนิคนี้ยังแบ่งย่อยออกได้เป็น

1. การกำหนดวัตถุประสงค์ภาคขยาย (Amplified Objectives) วิธีการนี้เป็นการพัฒนาต่อจากวัตถุประสงค์เชิงพฤติกรรม ทั้งนี้ในตัวของวัตถุประสงค์ภาคขยายจะมีการแจกแจงรายละเอียดเพิ่มขึ้น โดยบรรยายให้ทราบถึงสถานการณ์ของการสอบ วิธีการสร้างตัวเลือกและเกณฑ์ในการให้คะแนนกับวัตถุประสงค์เฉพาะนั้นๆ จากวัตถุประสงค์ที่เด่นชัดดังกล่าวจะทำให้สามารถสร้างข้อสอบในแนวเดียวกันออกมาได้เป็นจำนวนมาก ทั้งได้รูปแบบการเขียนข้อสอบที่มีความเป็นปรนัยสูง

2. การกำหนดเป็นโครงสร้างรูปแบบเฉพาะประจำข้อสอบ (Item Form) วิธีการนี้การสร้างข้อสอบจะเป็นการเขียนบรรยายถึงโครงสร้างหลักของคำถามที่จะใช้ เป็นหลักหรือแกน โดยที่บางส่วนของประโยคในโครงสร้างหลักจะถูกกำหนดให้มีลักษณะเป็นตัวแปร ลักษณะของการสร้างข้อสอบนั้น ทุกๆ ครั้งที่มีการตั้งคำถาม โครงสร้างหลักหรือแกนของประโยคจะยังคงสภาพเดิม แต่คำของตัวแปรจะถูกเปลี่ยนแปลง โดยวิธีการนี้จะทำให้ได้ข้อสอบเป็นจำนวนมากที่วัดพฤติกรรมถาวรไว้เหมือนกัน

(ข) วิธีการทางด้านภาษา (Linguistic-based Methods) เทคนิคนี้แบ่งย่อย ออกได้ เป็นดังนี้

1. ลักษณะการผูกประโยค วิธีการนี้เป็นการพยายามที่จะผูกโครงสร้างของ คำและวลีต่างๆหลายๆตอน เพื่อให้สามารถ เกิดประโยคขึ้นได้หลายๆประโยคที่มีข้อความและ ความหมายแตกต่างกันออกไป โดยใช้หลักการทางด้านไวยากรณ์และความหมายที่ได้ เป็นแนว ทرفในการสร้าง (ดูตัวอย่างใน Millman in Popham 1974 : 333)

2. การใช้หลักการเรียนรู้โนทัศน์ (Concept-based Testing) วิธีการ นี้ใช้แนวคิดเกี่ยวกับการจัดประเภท (Classification) ของวัตถุ หรือการใช้มโนทัศน์ใดๆ ในลักษณะที่เป็นเกณฑ์การแยก จัดประเภทสิ่งของ หรือปรากฏการณ์ที่มีลักษณะแตกต่างกันโดย อาศัยคุณสมบัติบางประการ การสร้างภาวะการสอบจะเป็นการสร้างสถานการณ์ที่ทำให้ผู้เรียน ต้องใช้หลักการ หรือมโนทัศน์ที่ใช้ในการจัดจำแนกสิ่ง ได้จากการเรียนรู้มา เป็นวิธีการในการตอบ

(ค) การใช้คอมพิวเตอร์ช่วยสร้างข้อสอบ (Computer-based Methods) เทคนิค นี้ยังแบ่งย่อยออกได้ เป็น

1. การสร้างธนาคารข้อสอบ วิธีการนี้ไม่แตกต่างไปจากแนวคิดของการสร้าง ธนาคารข้อสอบในแนวเดิม เพียงแต่เป็นการนำเอาข้อสอบที่สร้างขึ้นได้แล้ว ไปเก็บไว้ในตัวความ จำของเครื่องคอมพิวเตอร์ เวลาต้องการใช้ก็จะเป็นการสุ่มเลือกข้อสอบที่เก็บไว้ออกมาใช้ โดย อาจจะใช้คำสั่งให้ปรากฏออกมาทางหน้าจอทีวี หรือพิมพ์ออกมาทางเครื่องพิมพ์

2. การใช้เครื่องคอมพิวเตอร์ช่วยสร้างข้อสอบ แนวคิดนี้เป็นการเขียนโปรแกรม เพื่อให้เครื่องคอมพิวเตอร์ช่วยสร้างข้อสอบ โดยในตัวโปรแกรมคำสั่งสามารถที่จะสั่งให้เครื่อง- คอมพิวเตอร์สร้างคำถามและตัวเลือกที่ต่างกันออกมาได้ในแต่ละครั้งของการสร้าง ซึ่งลักษณะ เช่นนี้จะคล้ายคลึงกับวิธีการสร้างรูปแบบเฉพาะประจำข้อสอบ

ในทางปฏิบัติแล้ว เทคนิคต่างๆ เหล่านี้ยังคงมีข้อจำกัดอยู่มาก ลักษณะการเขียนบทความ นำเสนอแนวคิดต่างๆที่ เกี่ยวกับการสร้างข้อสอบ มีลักษณะที่เขียน เป็นหลักการกว้างๆทั่วไป เป็น ส่วนใหญ่ในบางประเด็น ผู้เขียนบทความเองก็ไม่อาจยกตัวอย่างให้ เกิดความเข้าใจอย่างชัดเจน ได้ ทั้งรอยดัดและฮาลาดิโนนาเองก็ยอมรับว่า เทคนิคต่างๆ เหล่านี้ยังมีข้อจำกัดอยู่มาก นอกจากนั้น เทคนิคการสร้างข้อสอบแต่ละอย่างก็จะมี ความเหมาะสมแตกต่างกันออกไปตามสถานการณ์ต่างๆ

(Roid and Haladyna 1980 : 309) เบิร์ก (Berk 1980 c : 52) โกวประเมิน

เทคนิคของการสร้างเหล่านี้ในแง่ความซัดเจนสำหรับผู้สร้างข้อสอบ ความสามารถในการนำ

ไม่ใช่ประกอบการเรียน ความสามารถที่จะใช้กับขอบเขตเนื้อหาที่กำหนด และความสะดวกในการใช้ พบว่าเทคนิคแต่ละอย่างก็มีข้อดีในบางด้าน และมีข้อจำกัดในบางด้านเช่นกัน

ในด้านการวิเคราะห์ค่าประจำข้อสอบ แนวคิดดั้งเดิมของการวิเคราะห์ข้อสอบในแบบอิงกลุ่ม คือการหาค่าอำนาจจำแนกและการหาค่าความยาก เมื่อมีการพัฒนาแบบสอบอิง เกณฑ์ขึ้นมาใช้ ข้อสอบจะได้รับการพิจารณาเน้นหนักไปถึงความตรงด้านเนื้อหา เป็นสิ่งสำคัญประการแรก เกลเซอร์ (Glaser 1963 : 520) ได้กล่าวไว้ว่า ในการสร้างแบบสอบอิง เกณฑ์เพื่อวัดผลสัมฤทธิ์ทางการเรียนนั้น ย่อมเป็นไปได้หลังจากการสอนแล้วจะมีข้อสอบที่มีค่าความยากเป็น 1 และไม่มีอำนาจจำแนกเลย แต่ข้อสอบนั้นๆก็ยังคงมีความจำเป็นที่จะต้องใช้สำหรับวัดผลสัมฤทธิ์ทางการเรียน เหตุนี้จึงเกิดมีแนวคิดคือไปว่า การวิเคราะห์ข้อสอบของแบบสอบอิง เกณฑ์น่าจะต้องใช้เกณฑ์หรือวิธีการที่แตกต่างออกไปจากการวิเคราะห์ในแนวเดิม เช่น ฮาลาดินา (Haladyna 1974 : 78) ได้ใช้การหาค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบร์ เรียลระหว่างคะแนนของข้อสอบแต่ละข้อกับคะแนนรวมจากแบบสอบเต็มฉบับ เมื่อนำเอาคะแนนที่ได้จากกลุ่มผู้สอบก่อนการสอนและคะแนนจากกลุ่มผู้สอบหลังการสอนมาวิเคราะห์ ซึ่งจากการวิเคราะห์ดังกล่าวนี้พบว่าทำให้ค่าอำนาจจำแนกของข้อสอบมีค่าเพิ่มขึ้น

นอกจากนั้นในระบบของการสอบแบบอิง เกณฑ์ ได้มีการพัฒนาดัชนีประจำข้อสอบขึ้นมาใช้โดยเฉพาะตัวหนึ่ง คือ "ดัชนีความไวในการสอน" (Instructional Sensivity Index ย่อ ISI) ซึ่งเป็นดัชนีชี้พัฒนาการของกลุ่มผู้เรียนในการทำข้อสอบข้อหนึ่งๆได้ถูกต้อง ฮาลาดินาและรอยด์ (Haladyna and Roid 1981) ได้ศึกษาเปรียบเทียบดัชนี ISI ซึ่งได้จากแนวคิดต่างๆ 4 แบบ คือ

1. แนวคิดของการสอบแบบอิง เกณฑ์ที่มีมาก่อน แนวคิดนี้ยังมีวิธีการคำนวณค่า ISI แยกย่อยออกเป็น 2 วิธี คือ

- ความแตกต่างระหว่างสัดส่วนของผู้ตอบถูกหลังการสอนกับก่อนการสอน (Pretest to Posttest Difference Index, PPDI) ซึ่งคอคซ์และวากัส (Cox and Vargas) เคยเสนอไว้ และเกรเฮน (Crehen 1974) ได้อ้างอิงถึงและนำมาใช้

- ค่าเปอร์เซ็นต์ที่เพิ่มขึ้น (The Percent of Possible Gain, PPG) คำนวณได้จากสูตร

$$PPG = \frac{PPDI}{1 - \text{Pretest Difficulty}}$$

2. แนวคิดของการวิเคราะห์ข้อสอบในแนวกังเคิม (Classical Test Theory) แนวคิดนี้ใช้การหาค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบเซรีระหว่างคะแนนจากแบบสอบเต็มฉบับกับคะแนนประจำข้อสอบที่เป็น 0 กับ 1 โดยกลุ่มตัวอย่างที่ใช้เกิดจากการรวมกลุ่มผู้สอบที่ได้รับการสอน และกลุ่มผู้สอบที่ยังไม่ได้รับการสอนเข้าด้วยกัน (Combined-samples point-biserial Correlation, COMPBI)

3. แนวคิดของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) แนวคิดนี้ใช้ความแตกต่างของค่าความยากประจำข้อสอบที่ได้รับการปรับให้เป็นค่าคะแนนมาตรฐานใต้พื้นโค้งปกติ (Normalized difference between item difficulty, ZDIFF) เป็นดัชนี

4. แนวคิดจากทฤษฎีของเบส์ ใช้การหาค่าดัชนี 3 ตัว คือ

- การหาค่าโอกาสที่ผู้ตอบข้อสอบข้อหนึ่งๆ จะเป็นผู้รอบรู้ที่แท้จริง (B1)
- การหาค่าโอกาสที่ผู้ตอบข้อสอบข้อหนึ่งๆ จะเป็นผู้รอบรู้ไม่จริง (B2)
- การหาค่าโอกาสที่จะตัดสินใจถูกต้องว่าผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ (B3)

ทั้งนี้ ดัชนีแต่ละตัวมีวิธีการคำนวณ คือ

$$B1 = \frac{(POSTDIFF) (COMDIFF)}{(POSTDIFF) (COMDIFF) + (PREDIFF) (1-COMDIFF)}$$

$$B2 = \frac{(1-PREDIFF) (1-COMDIFF)}{(1-PREDIFF) (1-COMDIFF) + (1-POSTDIFF) (COMDIFF)}$$

$$B3 = (POSTDIFF-COMDIFF) + (1-PREDIFF) + (COMDIFF-POSTDIFF)$$

เมื่อ	POSTDIFF	คือค่าความยากของข้อสอบหลังการสอน
	PREDIFF	คือค่าความยากของข้อสอบก่อนการสอน
	COMDIFF	คือค่าความยากของข้อสอบเมื่อนำผลการสอบก่อนและหลังการสอนมารวมกัน

จากการศึกษาพบว่า ค่ามัธยฐานของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างดัชนีต่างๆ มีสิ่งที่น่าสนใจคือ ค่ามัธยฐานของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างดัชนี PPDI กับดัชนี PPG, COMPBI, ZDIFF, B1, B2, B3, POSTDIFF, PREPBI, POSTPBI มีค่าเป็น 0.71,

0.79, 0.87, 0.38, 0.63, 0.59, 0.20 และ 0.30 ตามลำดับ (PREPBI และ POSTPBI คือค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยน์ไบร์เรียลระหว่างคะแนนจากแบบสอบเต็มฉบับก่อนและหลังการสอนกับคะแนนของข้อสอบแต่ละข้อ) ฮาลาไคนาและรอยด์ไค์เสนอว่าดัชนี ISI ที่มีประสิทธิภาพสูงและคำนวณได้ง่ายตัวหนึ่งคือ PPDI แต่ทั้งนี้ฮาลาไคนาและรอยด์ไค์กล่าวไว้ว่า การตัดสินใจเลือกข้อสอบในการสร้างแบบสอบอิงเกณฑ์นั้นไม่ควรจะพิจารณาเฉพาะค่าประจำข้อสอบเท่านั้น แต่จะต้องพิจารณาจากองค์ประกอบที่สำคัญอื่นๆด้วย โดยเฉพาะความสัมพันธ์ระหว่างเนื้อหาที่ข้อสอบมุ่งวัดกับการจัดการเรียนการสอนจริง

1.3 แนวคิดของการกำหนดเกณฑ์มาตรฐาน

แฮมเบิลตันและโนวิก (Hambleton and Novick 1973 : 163) ได้สรุปแนวคิดของแบบสอบอิงเกณฑ์ไว้ว่า เป็นแบบสอบที่มีการกำหนดเกณฑ์มาตรฐานหรือจุดตัด (Cut-off Score) ขึ้นไว้ ทั้งนี้เนื่องจากแบบสอบอิงเกณฑ์ประกอบด้วยข้อสอบที่สุ่มขึ้นมาจากรายชื่อจำนวนข้อสอบทั้งหมดภายในขอบเขตของเนื้อหาที่กำหนดขึ้น สัดส่วนจำนวนข้อสอบที่ผู้สอบทำได้คือจำนวนข้อสอบทั้งหมดจะแทนด้วยสัญลักษณ์ $\%$ โดยทางทฤษฎีค่า $\%$ จะถือว่าเป็นค่าประมาณของความรู้ที่แท้จริงของผู้สอบในขอบเขตของเนื้อหาที่กำหนด หรือมีความหมายว่าค่า $\%$ คือจำนวนข้อสอบทั้งหมดในขอบเขตของเนื้อหาที่ผู้สอบทำได้คือจำนวนประชากรข้อสอบทั้งหมด ถ้าหากผู้สอบได้ทำข้อสอบทุกข้อในขอบเขตนั้นๆ จุดตัดที่เป็นเกณฑ์มาตรฐานจะแทนด้วยสัญลักษณ์ $\%$ จุดตัดดังกล่าวนี้จะถือเป็นจุดแบ่งผู้สอบออกเป็นสองส่วน คือ ผู้สอบที่ได้คะแนนสูงกว่าเกณฑ์มาตรฐาน ($\% > \%$) ซึ่งจะถือว่าเป็นผู้รอบรู้ (Master) และผู้ที่ได้คะแนนต่ำกว่าเกณฑ์มาตรฐาน ($\% < \%$) ซึ่งจะถือว่าเป็นผู้ไม่รอบรู้ (Nonmaster)

ในทางปฏิบัติ เกณฑ์มาตรฐานหรือจุดตัด หมายถึง สัดส่วนจำนวนข้อสอบที่ผู้สอบทำได้คือจำนวนข้อสอบทั้งหมดในแบบสอบ ทั้งนี้ถ้าผู้สอบคนใดทำแบบสอบได้ถูกต้อง โดยมีสัดส่วนข้อสอบที่ทำถูกต้องจำนวนข้อสอบทั้งหมดในแบบสอบสูงกว่าสัดส่วนที่กำหนด จะถือว่าผู้สอบเป็นผู้รอบรู้ และถ้าผู้สอบทำแบบสอบได้ถูกต้อง โดยมีสัดส่วนข้อสอบที่ทำถูกต้องจำนวนข้อสอบทั้งหมด จะถือว่าผู้สอบเป็นผู้ไม่รอบรู้

การกำหนดเกณฑ์มาตรฐานนับได้ว่าเป็นส่วนสำคัญของแบบสอบอิงเกณฑ์ โดยมีผู้นำเสนอแนวคิด เทคนิค และวิธีการต่างๆในการกำหนดเกณฑ์มาตรฐานไว้เป็นจำนวนมาก ในระยะแรกๆแนวคิดต่างๆค่อนข้างจะกระจัดกระจาย และมีความแตกต่างกันมาก แต่ต่อมาแนวคิดต่างๆก็ค่อยๆ

เกิดการหลอมนตัวเอง เข้า เป็นหมวดหมู่และมีความเป็นระบบมากขึ้น โดยเฉพาะหลังช่วงปี 1980 ในที่นี้ใคร่ขอนำเสนอประเด็นต่างๆที่สำคัญ โดยยึดเอาลำดับขั้นของพัฒนาการในเรื่องนี้ตามเวลาที่ผ่านไปเป็นหลักดังนี้

มิลล์แมน (Millman 1973) ได้สรุปแนวคิดของการกำหนดเกณฑ์มาตรฐานไว้ว่า ควรจะคำนึงถึงสิ่งต่างๆต่อไปนี้

1. ระบุความสามารถในการทำแบบสอบของคนอื่นๆ (Performance of Others) โดยเฉพาะผลการสอบของกลุ่มที่ผ่านการยอมรับแล้ว ย่อมนำมาใช้เป็นเกณฑ์ในการพิจารณากำหนดเกณฑ์มาตรฐานได้

2. เนื้อหาสาระของข้อสอบในแบบสอบ (Item Content) เนื้อหาของข้อสอบเป็นส่วนสำคัญส่วนหนึ่งในการที่จะพิจารณากำหนดเกณฑ์มาตรฐาน ข้อสอบที่วัดในเนื้อหาที่มีความสำคัญย่อมจะต้องให้ผู้สอบทำได้ในสัดส่วนที่สูง แต่ในส่วนของเนื้อหาที่ไม่สำคัญนัก หรือมีความยากเกินไป เกณฑ์มาตรฐานก็ควรจะลดค่าลง

3. ผลของการศึกษาที่จะตามมาจากการกำหนดเกณฑ์มาตรฐาน (Educational Consequences) ถ้าหากว่ากำหนดเกณฑ์มาตรฐานค่าเกินไปก็อาจจะเกิดผลเสีย คือทำให้ผู้เรียนไม่สามารถเรียนบทเรียนที่สูงขึ้นไปได้ แต่ถ้ากำหนดเกณฑ์มาตรฐานสูงเกินไป ก็จะมีผลทำให้มีผู้ไม่รอบรู้ตกค้างเป็นจำนวนมาก ซึ่งอาจจะมีผลกระทบในขั้นอื่น เช่น จากผลการวิจัยของบลอค (Block อ้างถึงใน Glass 1978 : 256) พบว่าในทางพุทธศาสนามีความเป็นไปได้ที่จะสอนให้ผู้เรียนเกิดการเรียนรู้นทำแบบสอบได้ถูกต้องถึงร้อยละ 95 แต่จะมีผลเสียคือทำให้ทัศนคติในทางบวกต่อวิชาที่เรียนลดลง

4. ผลทางด้านจิตวิทยาและสภาพทางการเงิน (Psychological and Financial Cost) การกำหนดเกณฑ์มาตรฐานควรจะไต่คำนึงถึงผลทางด้านจิตใจที่เกิดขึ้นกับผู้สอบ เช่น แรงจูงใจ ความเบื่อหน่าย การทำลายอัทธมในทัศน และรวมทั้งค่าใช้จ่ายในการสอนซ่อมเสริมให้กับผู้เรียนที่สอบไม่ผ่าน เกณฑ์มาตรฐาน

5. ความคลาดเคลื่อนในการเดาและการสุ่มข้อสอบ (Error due to Chance and Item Sampling) การกำหนดเกณฑ์มาตรฐานควรจะไต่คำนึงถึงโอกาสที่ผู้สอบจะเดาถูกและรวมทั้งลักษณะการสุ่มข้อสอบจากประชากรข้อสอบทั้งหมดในขอบเขตเนื้อหาที่กำหนดขึ้นด้วย

สวามินาธาน, แฮมเบิลตัน และ แอลจินา (Swaminathan, Hambleton and Algina 1975) ได้ประยุกต์วิธีการของเบส์มาใช้เป็นมาตรการในการกำหนดเกณฑ์มาตรฐาน โดยใช้แนวคิดการจัดความสัมพันธ์ระหว่างการแบ่งผู้สอบออกเป็นผู้รอบรู้และผู้ไม่รอบรู้ กับการแบ่งผู้สอบในส่วนที่เกี่ยวข้องกับ โอกาสในการตัดสินใจถูกและผิดในการให้ผู้สอบสอบผ่านและสอบไม่ผ่าน ซึ่งสามารถจัดประเภทได้ดังนี้

ตารางที่ 3 ความสัมพันธ์ระหว่างภาวะการรอบรู้ของผู้สอบกับภาวะความถูกต้องในการตัดสินใจในการประยุกต์วิธีการของเบส์เพื่อกำหนดเกณฑ์มาตรฐาน

ภาวะการตัดสินใจ	ภาวะของผู้สอบ	
	รอบรู้	ไม่รอบรู้
ให้ผ่าน	ตัดสินใจถูก	เกิดความผิดพลาดให้ น้ำหนักเป็น 1_{12}
ไม่ผ่าน	เกิดความผิดพลาด ให้น้ำหนักเป็น 1_{21}	ตัดสินใจถูก

จากตารางสรุปดังกล่าว ความผิดพลาดในการตัดสินใจอาจเกิดขึ้นได้ใน 2 ลักษณะ คือ

1. ถ้าผู้สอบไม่รอบรู้แต่ตัดสินใจให้ผ่าน ความผิดพลาดจะเป็น $1_{12} \text{Prob}(\tau \leq \tau_0)$
2. ถ้าผู้สอบเป็นผู้รอบรู้แต่ตัดสินใจให้ตก ความผิดพลาดจะเป็น $1_{21} \text{Prob}(\tau > \tau_0)$

แต่ทั้งนี้ในสภาพความเป็นจริง ไม่มีโอกาสที่จะทราบได้ว่า การตัดสินใจถูกหรือผิด ทั้งนี้เพราะไม่ทราบว่า τ ของผู้สอบแต่ละคน จึงถือเอาว่า

1. ให้ผู้สอบสอบผ่าน และถือว่าเป็นผู้รอบรู้ ถ้าหากว่า

$$1_{21} \text{Prob}(\tau \leq \tau_0) < 1_{12} \text{Prob}(\tau > \tau_0)$$

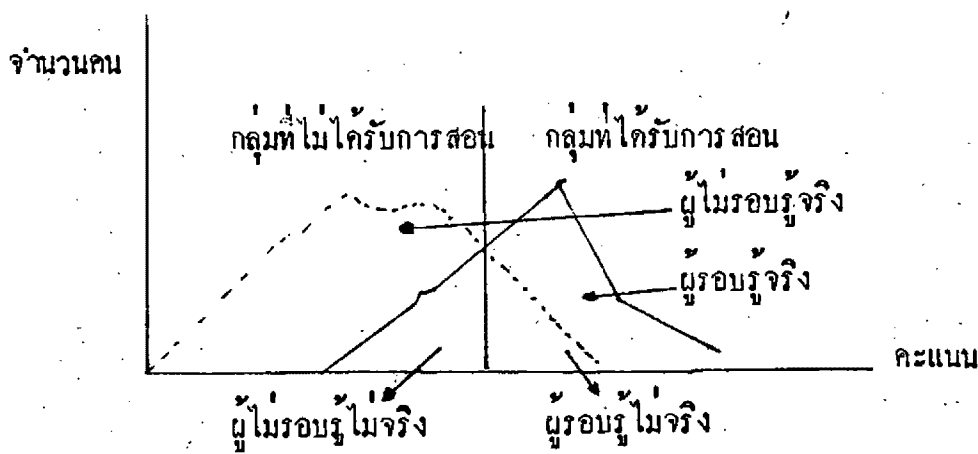
2. ให้อุสอบสอบตก และถือว่าเป็นผู้ไม่รอบรู้ ถ้าหากว่า

$$1_{21} \text{Prob}(T > T_c) < 1_{12} \text{Prob}(T \leq T_c)$$

วิธีการของเบย์มีข้อคืออยู่สองประการ คือ ประการแรกในการกำหนดเกณฑ์มาตรฐาน จะมีการนำเอาข้อมูลเกี่ยวกับระดับความสามารถของผู้สอบก่อนการสอบ เข้ามารวมพิจารณาด้วย และอีกประการหนึ่ง คือ มีการคำนึงถึงสัดส่วนค่าน้ำหนักของความผิดพลาดในแต่ละประเภท คือ การให้ผู้ออบรู้สอบตก และผู้ไม่รอบรู้สอบผ่าน ว่า ควรจะมีค่าน้ำหนักเป็นสัดส่วนเท่าใด

เบิร์ก (Berk 1976 : 6) ได้เสนอวิธีการกำหนดเกณฑ์มาตรฐานโดยใช้ลักษณะ การแจกแจงของคะแนนการสอบระหว่างกลุ่มที่ได้รับการสอน และกลุ่มที่ไม่ได้รับการสอน ทั้งนี้

แผนภูมิที่ 4 แนวคิดการกำหนดเกณฑ์มาตรฐานจากกลุ่มที่ได้รับการสอนและกลุ่มที่ไม่ได้รับการสอนของเบิร์ก



จากเส้นกราฟในแผนภูมิดังกล่าว การกำหนดเกณฑ์มาตรฐานควรจะกำหนดตรงจุดที่ทำให้เกิดความคลาดเคลื่อนในการจัดประเภทบุคคลคือ เกิดผู้รอบรู้ไม่จริง และผู้ไม่รอบรู้ไม่จริง เท่าๆกัน

ทั้งเบิร์กได้อธิบายไว้ด้วยว่า การกำหนดเกณฑ์มาตรฐานนั้นควรจะคำนึงถึงประเด็นต่างๆต่อไปนี้ประกอบ คือ

1. วัตถุประสงค์หรือคุณลักษณะที่ต้องการวัด
2. การดำเนินการ สอนตามวัตถุประสงค์ที่กำหนด
3. คาสถิติและลักษณะของ ข้อสอบในแบบสอบ
4. ความยาวของแบบสอบ
5. การประมาณค่าความสูญเสียและค่าที่ไ้ของแต่ละบุคคลและกลุ่มบุคคลที่เข้าสอบ

ทอม่า กลาส (Glass 1978) ใ้คำแนะนำเกี่ยวกับความการกำหนด เกณฑ์มาตรฐาน เท่าที่รวบรวมและจัดประเภทไว้ได้ แต่ทั้งนี้ การอภิปรายของกลาสปรากฏออกมาในลักษณะที่ว่า กลาสไม่เห็นด้วยกับการมี เกณฑ์มาตรฐาน ไม่ว่าจะด้วยวิธีการใดๆ เกณฑ์มาตรฐาน เท่าที่กลาสรวบรวมได้ มีอยู่ 6 วิธี คือ

1. การพิจารณาความสามารถของบุคคลอื่น (Performance of others as a Criterion) แนวคิดที่กลาสเห็นว่า ไม่แตกต่างไปจากแนวคิดของการประเมินผลแบบอิงกลุ่ม เพราะยังคงเป็นการอ้าง โดยการนำเอาข้อมูลจากผู้สอบคนอื่นมาใช้ เป็นหลักในการพิจารณา กำหนด เกณฑ์มาตรฐาน

2. การนับถอยหลัง (Counting backward from 100%) แนวคิดนี้พัฒนามาจากความคิดที่ว่า ผู้สอบควรจะมีความรู้ในเรื่องนั้นๆ อย่างสมบูรณ์ แต่ในทางปฏิบัติย่อมเป็นไปไ้ยาก จึงไ้มีการลด เกณฑ์ลง เช่น อาจจะทำเกณฑ์มาตรฐานเป็น 80% หรือ 90% กลาสไ้แย้งว่า การลด เกณฑ์ลงมาถึงกล่าวเป็นการกระทำอย่าง ไม่มีหลักการ และผู้กำหนด เกณฑ์มาตรฐานแต่ละคนก็มี เกณฑ์ที่ต่างกันออกไป

3. การใช้ เกณฑ์ภายนอกเข้ามาช่วยในการตัดสินใจ (Bootstrapping on other Criterion Scores) เช่น การใช้คะแนนจากกลุ่มบุคคลที่ผ่านการสอบและทำงานนั้น ได้ก็มาใช้เป็นหลักในการกำหนด เกณฑ์มาตรฐาน วิธีการนี้กลาสไ้แย้งว่า จะใช้ไ้ก็ก็ต่อเมื่อตัว เกณฑ์ภายนอกกับคะแนนจากแบบสอบอิง เกณฑ์มีความสัมพันธ์กันอย่างสมบูรณ์ มิฉะนั้นแล้ว ก็จะมีเกิดความคลาดเคลื่อนขึ้นไ้ คือ จะเกิดปรากฏการณ์มีคนสอบผ่าน เกณฑ์ภายนอก แต่ไ้คะแนนจากแบบสอบต่ำกว่า เกณฑ์มาตรฐานที่กำหนด และมีคนสอบไม่ผ่าน เกณฑ์ภายนอก แต่สอบผ่าน เกณฑ์มาตรฐานของแบบสอบ

4. การกำหนด สมรรถภาพต่ำสุด (Judging Minimal Competence) วิธีนี้ใช้ผู้เชี่ยวชาญเป็นผู้ร่วมในการกำหนด เกณฑ์มาตรฐานต่ำสุดที่ยอมรับได้ เช่น วิธีการกำหนด เกณฑ์



มาตรฐานของ อีเบล, แองกอพฟ์ และ นีเคลสกี

วิธีการกำหนด เกณฑ์มาตรฐานของอีเบล จะมีการคำนึงถึงเงื่อนไขต่างๆ คือ

1. ข้อสอบแต่ละข้อในแบบสอบอิง เกณฑ์นั้นมีความเกี่ยวเนื่องกับโคเมนที่ของการวัดในระคับโค เมื่อแบ่งออกเป็น สำคัญมาก สำคัญ ยอมรับได้ และ ไม่น่าใจ
2. ข้อสอบแต่ละข้อมีค่าความยากในความเห็นของผู้เชี่ยวชาญในระคับโค โดยคิดเป็นสัดส่วนว่า ถ้ามีคนเข้าสอบ 100 คน จะมีคนตอบถูกประมาณกี่คน

จากระคับความสำคัญของข้อสอบแต่ละข้อ และค่าความยากที่ผู้เชี่ยวชาญพิจารณาสามารถนำมากำหนด เป็น เกณฑ์มาตรฐานชั้นค่าได้ดังนี้

$$\text{เกณฑ์มาตรฐาน} = \frac{\sum (\text{ค่าความยากของข้อสอบในแต่ละระคับความสำคัญ} \times \text{จำนวนข้อสอบในระคับความสำคัญนั้นๆ})}{\text{จำนวนข้อสอบทั้งหมดในแบบสอบ}}$$

วิธีการกำหนด เกณฑ์มาตรฐานของแองกอพฟ์ เป็นวิธีการให้ผู้เชี่ยวชาญพิจารณา ข้อสอบแต่ละข้อ ผู้ที่มีความรู้ค่าสุดที่ยอมรับให้สอบผ่านได้ควรจะมีโอกาสตอบถูกเท่าใด ผลรวมของโอกาสที่จะตอบถูกจากข้อสอบทุกๆข้อ หากคูณจำนวนข้อสอบทั้งหมด ก็จะเป็น เกณฑ์มาตรฐานของแบบสอบอิง เกณฑ์ฉบับนั้น

วิธีการกำหนด เกณฑ์มาตรฐานของนีเคลสกี วิธีการนี้ใช้การให้ครูผู้สอนพิจารณาว่า ถ้านักเรียนคนหนึ่งที่มีความสามารถในระคับคาบเส้น คือ อยู่ระหว่าง เกรด ก ถึง เกรด จ ใ้ก่อนข้อคำถามหนึ่งๆแล้ว จะมีตัวเลือกใดบ้างที่นักเรียนดังกล่าวจะสามารถรู้ได้แน่ๆว่าเป็น คำลวงอย่างแน่นอน โดยวิธีการนี้ โอกาสที่ผู้สอบซึ่งมีความสามารถในระคับคาบเส้นจะตอบข้อสอบข้อหนึ่งๆถูกจะมีค่าเท่ากับ เศษหนึ่ง หกร้อยส่วน คือจำนวนตัวเลือกหลังจากที่ลบจำนวนคำลวงที่ผู้สอบในระคับคาบเส้นรู้ว่าผิดและไม่เลือกอย่างแน่นอน คะแนนค่าสุดของแบบสอบทั้งฉบับ ซึ่งผู้สอบที่มีความสามารถในระคับคาบเส้นจะทำได้ ก็คือผลรวมของโอกาสที่ผู้สอบในระคับคาบเส้นจะตอบข้อสอบแต่ละข้อถูก การกำหนด เกณฑ์มาตรฐานของแบบสอบโดยวิธีการนี้จะทำได้จาก

สูตร

$$\text{เกณฑ์มาตรฐาน} = M_{FD} + k \cdot \sigma_{FD}$$

- เมื่อ M_{FD} คือคะแนนเฉลี่ยที่ได้จากการกำหนดของครูผู้สอนทุกคน
- k คือค่าความคงที่ในลักษณะของคะแนนมาตรฐาน
- C_{FD} คือค่าส่วนเบี่ยงเบนมาตรฐานจากการตัดสินของครูผู้สอนแต่ละคน

วิธีการดังกล่าวนี้ กลาสโกล์แย้งว่ามีปัญหาหลายประการ เช่น ปัญหาการตัดสินใจได้อย่างมีเหตุผลในทางจิตวิทยาของผู้เชี่ยวชาญหรือครูผู้สอน ปัญหาความคงที่และความตรงในการตัดสินใจของผู้เชี่ยวชาญและครูผู้สอน ตลอดจนทั้งปัญหาการแปลความหมายของคำว่า "เกณฑ์มาตรฐานขั้นต่ำ"

5. การใช้ทฤษฎีการตัดสินใจประกอบ (Decision Theoretic Approach) วิธีการนี้มุ่งหาวิธีการทางสถิติ เพื่อนำมาใช้เป็นวิธีการในการกำหนดเกณฑ์มาตรฐาน เช่น การกำหนดเกณฑ์มาตรฐานที่จะทำให้เกิดความคลาดเคลื่อนในการจัดประเภทบุคคลตามระดับการรอบรู้ต่ำสุด วิธีการนี้กลาสโกล์เห็นว่าเป็นเพียงการกำหนดหลักการในการคำนวณเท่านั้น แต่การกำหนดค่าพื้นฐานต่างที่จะแทนลงไปในสมการทางคณิตศาสตร์ยังคงได้มาจากการตัดสินใจพิจารณาที่กำหนดขึ้น ซึ่งย่อมเป็นไปได้ว่า ผู้ที่กำหนดแต่ละคนย่อมสามารถที่จะกำหนดค่าพื้นฐานที่จะแทนลงไปในการได้แตกต่างกัน

6. การใช้วิธีการวิจัยเชิงปฏิบัติการ (Operations Research Method) วิธีการนี้เป็นการพยายามนำเอาคุณค่าภายนอกที่เหมาะสมสัมพันธ์กับคะแนนที่ได้จากแบบสอบ เช่น การพิจารณาถึงทัศนคติต่อวิชาที่เรียน ซึ่งจะผันแปรไปตามเกณฑ์มาตรฐานของแบบสอบ ดังเช่นจากผลการวิจัยของบลอคที่พบว่า ถ้ากำหนดเกณฑ์มาตรฐานสูงเกินไป แม้จะมีข้อดีในแง่ของพุทธนิสัย แต่จะมีผลตรงข้ามในทางจิตนิสัย กลาสโกล์แย้งว่า วิธีการนี้ไม่สามารถระบุได้ว่าบุคคลในลักษณะใด ในการเรียนวิชาใด จะมีเงื่อนไขอะไรบางอย่างที่จะนำมาใช้เป็นเกณฑ์ในการกำหนดเกณฑ์มาตรฐานของแบบสอบได้อย่างเหมาะสม

แฮมเบิลตัน (Hambleton 1978) ได้โต้แย้งกลับถึงแนวคิดของกลาสโกล์ที่กล่าวไว้ข้างต้น โดยแฮมเบิลตันมีความเห็นว่า การที่จะต้องมีเกณฑ์มาตรฐานยังคงเป็นสิ่งที่มีความจำเป็นในชีวิตประจำวันก็มีการกำหนดและการใช้เกณฑ์มาตรฐานในลักษณะของการกำหนดขึ้น เช่น เกณฑ์มาตรฐานของสุขภาพ เกณฑ์มาตรฐานความปลอดภัยบนทางหลวง การกำหนดเกณฑ์มาตรฐานอาจจะเป็นสิ่งที่มีปัญหาในการตัดสินความสามารถขั้นต่ำสุดของบุคคล เพื่อให้ประกาศนียบัตร

หรือการประเมินโครงการต่างๆ แต่ในระดับชั้นเรียนแล้วครูผู้สอนน่าจะเป็นผู้ที่มีความสามารถ
ร่วมกันพิจารณาว่าหนดชั้นได้

จากการที่มีข้อโต้แย้งความคิดที่กระจัดกระจายออกไปมาก เบิร์กจึงได้กล่าวไว้
(Berk 1980 a : 101) ในบทความก่อนที่จะอภิปรายถึงวิธีการต่างๆในการกำหนดเกณฑ์
มาตรฐานว่า มาตรการของการกำหนดเกณฑ์มาตรฐานทุกวิธี เป็นการลงข้อคัดค้านใจกำหนดชั้น
เหตุของการกำหนดเกณฑ์มาตรฐานจึงควรพิจารณาให้รอบคอบถึงปัจจัยต่างๆ คือ

1. ความสำคัญของการคัดสรรใจ
2. เวลาที่ทองการไว้เพื่อการคัดสรรใจ
3. ทรัพยากรบุคคลและงบประมาณที่มี
4. ความสามารถของผู้คัดสรร

นอกจากนั้น ลิฟวิงสตัน และ ซีกี (Livingston and Zieky 1982) ได้อภิปราย
ไว้อย่างชัดเจนว่า การกำหนดเกณฑ์มาตรฐาน คือ การพยายามตอบคำถามว่า "เท่าไรจึงจะถือว่าพอ"
หรือตอบคำถามว่า "คือเท่าไรจึงจะถือว่าดีพอ" ซึ่งการกำหนดเกณฑ์มาตรฐานทุกอย่าง
เป็นเรื่องของความเป็นอันหนึ่งอันเดียว ความสำคัญอยู่ที่ว่า เมื่อมีการกำหนดเกณฑ์มาตรฐานชั้น
มาแล้ว การคัดสรรใจต่างๆโดยการยึดเกณฑ์มาตรฐานจะสามารถทำได้อย่างมีความเป็นปรนัย

จากการที่ไทม์นิกิวการเสนอข้อเขียนของคนเกี่ยวกับวิธีการกำหนดเกณฑ์มาตรฐาน
ขึ้นมามาก ในระยะต่อมาจึงได้มีการจัดประเภทและแยกหมวดหมู่แนวคิดของการกำหนดเกณฑ์
มาตรฐาน เกณฑ์ที่เกณฑ์หนึ่งซึ่งใช้เป็นหลักในการจัดแบ่งประเภทวิธีการกำหนดเกณฑ์มาตรฐาน
คือ เกณฑ์การใช้ข้อมูลเพื่อประกอบการคัดสรรใจกำหนดเกณฑ์มาตรฐาน ซึ่งแบ่งออกได้เป็น 2
วิธีการ คือ

1. การกำหนดเกณฑ์มาตรฐานโดยใช้ข้อมูลจากผู้เชี่ยวชาญเพียงประการเดียว
2. การกำหนดเกณฑ์มาตรฐานโดยใช้ข้อมูลเชิงสัมพันธ์ (Relative) จาก

ภายนอกเข้ามาเป็นส่วนประกอบในการพิจารณาเกณฑ์มาตรฐาน

เมื่อเริ่มเห็นความแตกต่างในวิธีการกำหนดเกณฑ์มาตรฐานทั้งสองวิธีดังกล่าวแล้วชัด
เจนขึ้น ก็ได้เกิดความพยายามกำหนดเกณฑ์มาตรฐานในวิธีการใหม่ คือ การนำเอาแนวคิดทั้ง
สองแนวเข้ามาผสมผสานเข้าด้วยกัน เช่น ครอสส์, อิมพารา, ฟรารี และ เจเกอร์ (Cross,
Impara, Frary and Jaeger 1984) ได้ใช้วิธีการกำหนดเกณฑ์มาตรฐานของแองกอฟฟ์,

นี่เคลสกี และ เจเกอร์ เป็นวิธีการในการกำหนดเกณฑ์มาตรฐานให้กับแบบสอบ (The National Teacher Examination" ของ "The Virginia Department of Education" โดยในขั้นตอนของการดำเนินงาน ได้นำเอาแนวคิดของการอาศัยข้อมูลจากภายนอกเข้ามาเป็นตัวร่วมพิจารณา คือ

1. ในรอบแรก ผู้ตัดสินจะได้รับข้อมูลเฉพาะแบบสอบข้อคำถามพิจารณา โดยไม่มีข้อมูลอื่นใดประกอบ

2. ในรอบที่สอง ผู้ตัดสินจะได้รับแบบสอบเฉพาะข้อคำถามพิจารณา แต่จะมีข้อมูลประกอบคือค่าความยากของข้อสอบแต่ละข้อ

3. ในรอบสุดท้าย ผู้ตัดสินจะได้รับการขอร้องให้พิจารณาอีกครั้งหนึ่ง โดยพิจารณาเฉพาะข้อคำถาม และมีข้อมูลประกอบ คือ ข้อมูลที่ได้จากการตัดสินของผู้ตัดสินคนอื่นๆในรอบแรก และข้อมูลความถี่ของคะแนนที่ผู้สอบทำได้จากการเล่นแบบสอบไปใช้

ผลการศึกษาพบว่า เมื่อมีการปรับคะแนนที่ได้จากผู้ตัดสินให้เป็นคะแนนจากแบบสอบเต็มฉบับ ผลการวิเคราะห์ความแปรปรวนได้ข้อสรุปว่า วิธีการกำหนดเกณฑ์มาตรฐานต่างกัน และผลการตัดสินใจในแต่ละรอบซึ่งมีข้อมูลจากภายนอกต่างกัน มีความสัมพันธ์เชิงเดียว เป็นเกณฑ์มาตรฐานแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001

เมื่อพิจารณาจากผลการวิจัยและข้ออภิปรายต่างๆแล้ว จะเห็นว่าวิธีการกำหนดเกณฑ์มาตรฐานในปัจจุบันยังไม่มีวิธีที่ดีที่สุด หรือมีสูตรสำเร็จในการกำหนดเกณฑ์มาตรฐาน แต่ทั้งนี้คงไม่ได้หมายความว่าเรามีเกณฑ์มาตรฐานประกอบการพิจารณาตัดสินใจจักประเภทบุคคลตามระดับการรอบรู้เป็นเรื่องไร้สาระ เพียงแค่ในการกำหนดเกณฑ์มาตรฐานนั้นคงจะต้องทำเป็นเฉพาะกรณีไป โดยคำนึงถึงข้อมูลในหลายๆด้าน และเมื่อได้เกณฑ์มาตรฐานมาแล้วก็ควรจะใช้เกณฑ์นั้นอย่างมีความเป็นปรนัย

1.4 ความตรงของแบบสอบอิงเกณฑ์

แม้ความตรงจะเป็นคุณสมบัติที่สำคัญของแบบสอบทุกประเภท แต่จากการค้นคว้าในเรื่องความตรงของแบบสอบอิงเกณฑ์ พบว่ายังมีการศึกษาไม่มากนัก ที่มีปรากฏชัดแจ้งก็คือ ข้อเขียนของนิตโกและแอมเบิลตัน ซึ่งมีประเด็นที่สำคัญดังนี้

นิตโก (Nitko 1980 : 478 - 479) ได้อภิปรายถึงความตรงประเภทต่างๆ ของแบบสอบอิงเกณฑ์ไว้ 3 ประเด็น คือ

1. ความตรงตามเนื้อหา (Content Validity) ความตรงตามเนื้อหาของแบบสอบอิงเกณฑ์ มีความหมายเน้นหนักถึงความเป็นตัวแทนของมวลความรู้ทั้งหมดในขอบเขตของเนื้อหาที่กำหนด ซึ่งถ้ามีการกำหนดเนื้อหาอย่างเด่นชัดแล้ว การที่จะสร้างแบบสอบให้มีความตรงตามเนื้อหาที่สามารถทำได้ไม่ยากนัก แต่ทั้งนี้จะต้องมีการวางแผนการสร้างและสุ่มข้อสอบเป็นอย่างดี

2. ความตรงตามโครงสร้างทฤษฎี (Construct Validity) ลักษณะของแบบสอบอิงเกณฑ์บางส่วนตามการจักประเภทของนิทโก มีความจำเป็นที่จะต้องมีความตรงตามโครงสร้างทฤษฎี เช่น แนวคิดของการพัฒนาแบบสอบอิงเกณฑ์ที่ยึดเอาพัฒนาการของขั้นตอนการเรียนรู้ (Learning Hierachy) แบบสอบที่พัฒนามาจากแนวคิดในเรื่องใดเรื่องหนึ่ง ที่เชื่อว่ามีพัฒนาการของความคิดเป็นลำดับขั้นตอน นอกจากนั้นจากการตัดสินใจจักประเภทบุคคลตามระดับการรอบรู้ ถ้าหากแบบสอบมีความตรงจริงแล้วก็น่าจะตรวจสอบได้ในแง่ที่ว่า ผู้สอบที่เป็นผู้รอบรู้จะมีความแตกต่างจากผู้สอบที่เป็นผู้ไม่รอบรู้ในลักษณะบางประการ เช่น ผู้ที่รอบรู้ควรจะเรียนบทเรียนที่สูงขึ้นไปไ้ได้ง่ายกว่าหรือเร็วกว่าผู้สอบที่เป็นผู้ไม่รอบรู้ ซึ่งนิทโกกล่าวไว้ว่า ประเด็นต่างๆ เหล่านี้ยังไม่มีการศึกษาเท่าที่ควร

3. ความตรงอันเนื่องมาจากเกณฑ์มาตรฐานของแบบสอบ (Criterion-oriented Validity) แนวคิดในประเด็นนี้มุ่งไปยังจุดเน้นที่ว่า ถ้าแบบสอบอิงเกณฑ์ที่พัฒนาขึ้นกล่าวอ้างอิงว่าสามารถระบุสิ่งใดสิ่งหนึ่งได้ เช่น บ่งบอกจุดบกพร่องในการเรียนของผู้สอบ หรือการจักประเภทบุคคลออกได้ เป็นกลุ่มๆ ตามลำดับคุณภาพใดๆแล้ว ก็ควรจะมีมาตรการในการอ้างอิงหรือยืนยันไว้ว่าสิ่งที่บ่งบอกจากแบบสอบ เป็นสิ่งที่มีความถูกต้อง

แฮมเบิลตัน (Hambleton in Berk 1980 : 80 - 99) ได้อภิปรายถึงประเด็นต่างๆ เกี่ยวกับความตรงของแบบสอบอิงเกณฑ์ไว้เช่นกัน โดยมีสาระสำคัญที่ควรจะกล่าวถึงดังนี้

ความตรงตามเนื้อหา แฮมเบิลตันได้อ้างถึงงานวิจัยของ ไฮวี่ลี, แพคเตอร์สัน และเพค ที่ทำในปี 1968 ว่า เป็นแบบสอบอิงเกณฑ์ที่มุ่งเน้นความตรงตามเนื้อหาที่มีเงื่อนไขสำคัญ 2 ประการ คือ ประการแรก ข้อสอบทั้งหมดในขอบเขตของเนื้อหาที่กำหนดได้รับการเขียนไว้ล่วงหน้าและรูขอบเขตอย่างเด่นชัด อีกประการหนึ่งคือข้อสอบทั้งหมดในแบบสอบได้มาจากการสุ่ม

เพื่อให้ได้แบบสอบที่มีคุณภาพในแง่ของความตรงตามเนื้อหา แบบสอบจะต้องประกอบด้วยข้อสอบที่มีคุณภาพที่มาก่อน ข้อสอบที่มีคุณภาพที่นั้น พิจารณาได้จากคุณสมบัติ 3 ประการ คือ

1. มีความตรงของข้อสอบ (Item Validity) ซึ่งอาจจะทำได้ 2 วิธี คือ วิธีแรก ให้ผู้รู้ในเนื้อหาของข้อสอบพิจารณาตัดสิน ซึ่งแอมเบิลตันเห็นว่า เป็นวิธีการที่ดี ส่วนอีกวิธีหนึ่ง คือ การใช้ข้อมูลเชิงประจักษ์ ใ้แก่การหาค่าอำนาจจำแนกของข้อสอบตามแนวคิดเดิม วิธีการนี้ แอมเบิลตันเห็นว่าอาจจะมีปัญหาในทางปฏิบัติ เพราะอาจจะเป็นไปได้ที่ข้อสอบและคะแนนจากแบบสอบ คัมฉบบมีความแปรปรวนค่า

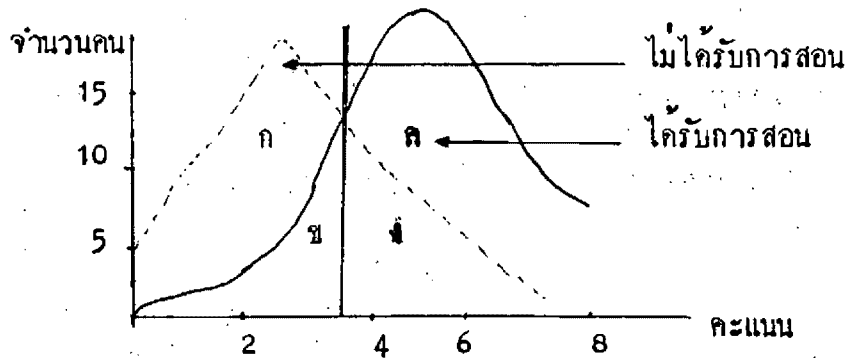
2. การตรวจสอบคุณภาพเชิงเทคนิคของการเขียนข้อสอบ (Technical Quality of Item Writing) ประเด็นนี้มุ่งเน้นถึงการเขียนข้อสอบให้มีคุณภาพในแง่ของหลักการเขียนข้อสอบโดยทั่วไป

3. ความเป็นตัวแทนของข้อสอบในแบบสอบ (Representativeness of the Test Item) จำนวนข้อสอบที่รวมกันเข้า เป็นแบบสอบ คัมฉบบจะต้องมีความเป็นตัวแทนของความรู้ในขอบเขตของเนื้อหาทั้งหมด

ความตรงตามเกณฑ์ของแบบสอบ (Criterion-Related Validity) แอมเบิลตันใช้คำนี้ในความหมาย เช่นเดียวกับคำว่า "ความตรงอันเนื่องมาจากเกณฑ์มาตรฐานของแบบสอบ" คือมุ่งเน้นประเด็นที่ว่า จากจุดเกณฑ์มาตรฐานที่กำหนดและใช้เป็นจุดแบ่งผู้สอบออกเป็นผู้รอบรู้และไม่รอบรู้นั้น กลุ่มผู้เรียนที่ได้รับการแบ่งออกไปน่าจะต้องมีความแตกต่างในเกณฑ์ภายนอกอื่นๆ เช่น การทำนายผลการเรียนในบทเรียนต่อไป ซึ่งเทคนิคหรือวิธีการที่จะใช้ในการศึกษาก็จะไม่แตกต่างไปจากวิธีการต่างๆ ที่มีใช้กันอยู่โดยทั่วไป

ความตรงในการตัดสินใจ (Decision Validity) ความตรงประเภทนี้จักเป็นความตรงเฉพาะชนิดหนึ่งของความตรงตามโครงสร้างทฤษฎี วิธีการตรวจสอบความตรงในการตัดสินใจของแบบสอบอิงเกณฑ์จะทำได้โดยการเปรียบเทียบการทำแบบสอบของกลุ่มที่ได้รับการสอนกับกลุ่ม ซึ่งอาจจะพิจารณาได้จากแผนภูมิดังนี้

แผนภูมิที่ 5 การพิจารณาความตรงในการตัดสินใจจากกลุ่มผู้สอบที่ได้รับการสอน และกลุ่มผู้สอบที่ไม่ได้รับการสอน



- เมื่อ ก คือกลุ่มผู้สอบที่ไม่ได้รับการสอนและเป็นผู้ไม่รอบรู้ เป็นการตัดสินใจที่ถูกต้อง
- ข คือกลุ่มผู้สอบที่ได้รับการสอน แต่เป็นผู้ไม่รอบรู้ เป็นการตัดสินใจที่ผิด
- ค คือกลุ่มผู้สอบที่ได้รับการสอนและเป็นผู้รอบรู้ เป็นการตัดสินใจที่ถูกต้อง
- ง คือกลุ่มผู้สอบที่ไม่ได้รับการสอน แต่เป็นผู้รอบรู้ เป็นการตัดสินใจที่ผิด

ซึ่งความตรงประเภทนี้ขึ้นอยู่กับปัจจัยต่างๆ คือ

1. คุณภาพของแบบสอบที่นำมาศึกษา
2. ความเหมาะสมของกลุ่มผู้สอบที่นำมาใช้ในการกำหนดเกณฑ์มาตรฐาน
3. ลักษณะของกลุ่มที่มุ่งจำแนก จัดประเภทระกับการรอบรู้
4. เกณฑ์มาตรฐานค่าสุดท้ายที่กำหนด

เกรเฮน (Crehen 1974) ได้อ้างถึงสูตรของคาร์เวอร์ที่ใช้ในการหาค่าความตรงของแบบสอบอิงเกณฑ์เมื่อแบ่งผู้สอบออกเป็น 4 กลุ่ม ดังตาราง

ตารางที่ 4 แนวคิดการหาค่าความตรงของแบบสอบอิงเกณฑ์โดยใช้สูตรของคาร์เวอร์

ผลการตัดสินใจ \ สภาวะการสอน	ไม่ได้รับการสอน	ได้รับการสอน
	รอบรู้	b
ไม่รอบรู้	c	d

ดังนั้น ความตรงในการตัดสินใจจากประเภทบุคคลตามระดับการรอบรู้ จะคำนวณได้

จากสูตร

$$\text{ค่าความตรง} = \frac{a + c}{a + b + c + d}$$

1.5 การหาค่าความเที่ยงของแบบสอบอิงเกณฑ์

เนื่องจากแบบสอบอิงเกณฑ์ไม่ไถ่มุ่งเน้นการจำแนกผู้สอบในกลุ่มเดียวกัน และทั้งยังเป็นไปไ้กว่า ผู้สอบอาจจะได้คะแนนแตกต่างกันไม่มากนัก ซึ่งกรณีเช่นนี้ย่อมมีแนวโน้มที่ว่าแบบสอบที่สร้างขึ้นมานั้นจะมีค่าความเที่ยงที่หาโดยยึดแนวคิดทฤษฎีดั้งเดิม (Classical Test Theory) ถ้า ทั้งที่ความเป็นจริงแล้วแบบสอบนั้นมีคุณภาพดี จึงได้มีผู้พยายามพัฒนาสูตรการหาค่าความเที่ยงซึ่งเหมาะสมที่จะใช้กับแบบสอบอิงเกณฑ์ขึ้นมาใหม่ แต่ทั้งนี้ไปแฟมและฮูเสค (Popham and Husek 1969 : 5) ก็ได้กล่าวไว้ว่า ถ้าหากว่าแบบสอบนั้นมีความแปรปรวนของคะแนนสูงพอ แนวคิดของการหาค่าความเที่ยงในแนวคิดทฤษฎีดั้งเดิมก็ยังคงสามารถนำมาใช้ได้ ผลของการพัฒนาสูตรในการหาค่าความเที่ยงที่ใช้กับแบบสอบอิงเกณฑ์โดยเฉพาะนั้น เบิร์ก (Berk 1980 a) ได้จัดกรวบรวมไว้ และแยกประเภทหรือกลุ่มแนวคิดไว้เป็น 3 กลุ่มด้วยกัน คือ

1. แนวคิดของการหาค่าความคงที่ในการจัดประเภทบุคคล (Reliability of Mastery Classification or Threshold Loss Function)

2. แนวคิดของการหาค่าความแตกต่างของคะแนนที่สอบได้กับคะแนนเกณฑ์มาตรฐาน (Reliability of Criterion-Referenced Test Score or Square Error Loss Function)

3. แนวคิดของการประมาณค่าคะแนนที่แท้จริงของผู้สอบ (Reliability of Domain Score Estimates or Domain Score Estimation)

โดยแนวคิดแต่ละประเภทมีรายละเอียดเฉพาะลงไป ดังนี้

1. แนวคิดของการหาค่าความคงที่ในการจัดประเภทบุคคล

หลักการสำคัญของแนวคิดนี้คือว่า ความแปรปรวนของแบบสอบถามซึ่งเกณฑ์ไม่มีความสำคัญ การหาค่าความเที่ยงของแบบสอบถามพิจารณาจากความคงที่ในการจัดประเภทบุคคลว่า เป็นผู้ที่มีความรอบรู้ในระดับใดเมื่อแยกเอาเกณฑ์มาตรฐานเป็นจุดแบ่ง ถ้าผลการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้มีความคงที่ไม่ว่าจะมีการวัดกี่ครั้งในเรื่องเดียวกัน ก็ถือว่าแบบสอบถามนั้นมีค่าความเที่ยงสูง โดยแฮมเบิลตันและโนวิก (Hambleton and Novick 1973) ได้เสนอให้ใช้แนวคิดของคาร์เวอริที่ว่า ความเที่ยงของแบบสอบถามจะหาได้จากความคงที่ในการจัดประเภทบุคคลตามระดับการรอบรู้โดยแยกเอาเกณฑ์มาตรฐานที่กำหนดเป็นจุดแบ่ง และใช้แบบสอบถามชุดหนึ่งในการทดสอบเพื่อหาค่าความเที่ยง โดยมีแนวการวิเคราะห์ข้อมูลที่ได้ดังนี้

ตารางที่ 5 การจำแนกข้อมูลเพื่อการหาค่าความเที่ยงของแบบสอบถามเกณฑ์ในแนวคิดการหาค่าความคงที่ในการจัดประเภทบุคคล

		ฟอร์ม ก	
		ผ่าน	ไม่ผ่าน
ฟอร์ม	แบบสอบ		
	ผ่าน	b	a
ข	ไม่ผ่าน	c	d

ดังนั้น ความเที่ยงจะหาได้จากสูตร

$$\hat{p} = \frac{b + d}{a + b + c + d}$$

ต่อมา สวามินาธาน, แฮมเบิลตัน และ แอลจินา (Swaminathan, Hambleton and Algina 1974) ได้เสนอแนวคิดการหาค่าความเที่ยงโดยใช้ดัชนีแคป้า (k) โดยมีสูตรในการคำนวณคือ

$$k = \frac{\hat{P}_o - P_c}{1 - P_c}$$

เมื่อ $\hat{P}_o = \sum_{k=1}^m P_{kk}$

$$\hat{P}_c = \sum_{k=1}^m P_{i \cdot} \times P_{\cdot i}$$

m คือจำนวนระดับการรอบรู้

โดยที่ P_c คือค่าของโอกาสหรือความบังเอิญที่จะจับบุคคลเข้ากลุ่มได้อย่างคงที่ถึงสองครั้ง การหาค่าดัชนีแคป้าสามารถที่จะใช้ได้ทั้งกรณีที่มีการจับประเภทบุคคลออกไปมากกว่าสองประเภท ทั้งยังสามารถที่จะหาค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) ของค่าแคป้าได้จากสูตร

$$s = \sqrt{\frac{\hat{P}_o (1 - \hat{P}_o)}{N (1 - \hat{P}_o)^2}}$$

n คือจำนวนผู้เข้าสอบทั้งหมด

ถ้าต้องการตรวจสอบว่า ค่าแคป้าที่คำนวณได้มีค่าสูงกว่า "0" อย่างมีนัยสำคัญทางสถิติหรือไม่ สามารถที่จะคำนวณได้จากสูตร

$$Z = \frac{\frac{\hat{k}}{\hat{P}_c}}{\sqrt{\frac{1 - \hat{P}_c}{N (1 - \hat{P}_c)^2}}}$$

(Cohen 1960 : 42 - 47)

อย่างไรก็ดี การหาค่าความเที่ยงจากสูตรดังกล่าวมีความจำเป็นที่จะต้องใช้แบบ
สอบคู่ขนาน หรือมีการทดสอบซ้ำ ซึ่งไม่สะดวกในทางปฏิบัติ จึงได้มีผู้พยายามพัฒนาสูตรขึ้นมา
ทดแทนแนวคิดนี้ โดยมุ่งให้สามารถหาค่าความเที่ยงจากแบบสอบโดยใช้การทดสอบเพียงครั้ง
เดียว บุคคลที่นำเสนอแนวคิดนี้ออกมาก็คือ สับโคเวียค (Subkoviak) ฮวน (Huynh)
และ มาร์แชลล์ (Marshall) โดยฮวนใช้หลักการของเบต้า ไปโนเนียล เป็นข้อตกลง
เบื้องต้น ส่วนสับโคเวียคและมาร์แชลล์ ใช้ข้อตกลงเบื้องต้นของการแจกแจงแบบทวินามเป็น
สำคัญ แนวคิดของการพัฒนาสูตรที่เข้าใจได้ง่ายแนวคิดหนึ่ง คือ วิธีการของสับโคเวียค ซึ่งมี
การพัฒนาสูตรตามลำดับขั้น ดังนี้

จากขั้นนี้ความสอดคล้องที่ว่า

$$P_o = \sum_{i=1}^k P_{ii}$$

เมื่อมีการดำเนินการทดสอบเพียงครั้งเดียว กำหนดให้ X และ X' เป็นคะแนน
ที่ได้จากแบบทดสอบที่เป็นคู่ขนาน ขั้นนี้ความสอดคล้องของคนที่ i ณ จุดเกณฑ์มาตรฐานที่กำหนด
จะเป็น

$$P_c^{(i)} = P(X_i \geq C, X'_i \geq C) + P(X_i < C, X'_i < C)$$

ขั้นนี้ความสอดคล้องของทุกคนก็จะเป็น

$$P_c = \sum_{i=1}^N P_c^{(i)}$$

เมื่อ N คือจำนวนผู้เข้าสอบทั้งหมด

ถ้ามีข้อตกลงเบื้องต้นว่า การกระจายของโอกาสที่คนที่ i ในแบบสอบทั้งสองจะ
ตอบ ถูก หรือ ผิด เป็นอิสระต่อกัน ก็จะได้ว่า

$$P_c^{(i)} = P(X_i \geq C) P(X'_i \geq C) + P(X_i < C) P(X'_i < C)$$

เมื่อมีข้อตกลงเบื้องต้นคือไปว่า การกระจายของคะแนนในแบบสอบทั้งสองฉบับ มีการกระจายแบบทวินาม ซึ่งข้อตกลงเบื้องต้นนี้จะเป็นจริงได้ ก็จะต้องมีข้อตกลงเบื้องต้นย่อยๆ คือ การให้คะแนนในข้อสอบแต่ละข้อเป็น 1 กับ 0 คะแนนที่ได้ในข้อสอบข้อต่างๆ มีความเป็นอิสระต่อกัน และโอกาสที่จะตอบข้อสอบข้อหนึ่งๆถูก มีค่าเท่ากันตลอด

จากข้อตกลงเบื้องต้นดังกล่าว ก็จะได้สมการต่อไปนี้ว่า

$$P_c^{(i)} = [P(X_i \geq c)]^2 + [1 - P(X_i \geq c)]^2$$

โดยที่

$$P(X_i \geq c) = \sum_{x_i=c}^n \binom{n}{x_i} P_i^{x_i} (1 - P_i)^{n - x_i}$$

ทั้งนี้ ค่า P_i จะเป็นค่าสัดส่วนของคะแนนที่สอบได้จากจำนวนข้อสอบทั้งหมด คือ n ข้อ โดยในการคำนวณจริง ค่า P_i จะเป็นค่าสัดส่วนที่แท้จริงของคะแนนของผู้สอบ ซึ่งจะได้จากการประมาณจากสูตรการแปลงคะแนนเชิงเส้นตรงของสมการที่มุ่งหาค่าคะแนนที่แท้จริงจากการสอบ โดยอาศัยแนวคิดของทฤษฎีคังเคิม คือ

$$P_i = \alpha_{21/x} \left(\frac{x_i}{n} \right) + (1 - \alpha_{21/x}) \left(\frac{M_x}{n} \right)$$

เมื่อ $\alpha_{21/x}$ คือเป็นค่าความเที่ยงในลักษณะของการหาค่าความคงที่ภายใน

สุดท้ายจะได้ค่าดัชนีความสอดคล้อง หรือค่าความเที่ยงตามวิธีการของสับโคเวียค

คือ

$$P_c = \frac{\sum_{i=1}^n P_c^{(i)}}{N}$$

2. แนวคิดของการหาค่าความแตกต่างของคะแนนที่สอบได้กับคะแนน เกณฑ์มาตรฐาน

แนวคิดนี้ถือว่า ระยะห่างของคะแนนที่สอบทำได้กับ เกณฑ์มาตรฐานที่กำหนด มีความสำคัญหรือเป็นข้อมูลที่คองนำไปนำมาช่วยประกอบในการตัดสินใจ จึงควรจะนำเข้ามาร่วมพิจารณาในการหาค่าความเที่ยงของแบบสอบ ส่วนแนวคิดอื่น ๆ ยังคงยึดแนวคิดของทฤษฎีคังเกิมอยู่ สูตรแรกที่ปรากฏตามแนวคิดนี้ คือสูตรของลิฟวิงสตัน (Livingston 1972) ซึ่งปรากฏอยู่ในรูป

$$K^2(X, T) = \frac{\sigma_T^2 + (U - C)^2}{\sigma_0^2 + (U - C)^2}$$

เมื่อ $K^2(X, T)$ คือค่าความเที่ยงของแบบสอบ

σ_T^2 คือค่าความแปรปรวนของคะแนนที่แท้จริง ซึ่งจะคำนวณได้จากสูตร

$$\sigma_T^2 = \infty \sigma_0^2$$

เมื่อ ∞ คือค่าความเที่ยงของแบบสอบในแนวทฤษฎีคังเกิม

σ_0^2 คือค่าความแปรปรวนของคะแนนที่ได้จากแบบสอบ

U คือค่ามัธยฐานเลขคณิตของกลุ่มผู้เข้าสอบ

C คือค่า เกณฑ์มาตรฐานที่กำหนด

ทั้งนี้ แนวคิดของการพัฒนาสูตรนั้น ลิฟวิงสตัน เริ่มกันจากแนวคิด คังเกิมที่นิยามว่า

$$S_x^2 = \frac{\sum (x - \bar{x})^2}{N}$$

เนื่องจากแบบสอบอิง เกณฑ์ยึด เอาจาก เกณฑ์มาตรฐานเป็นหลัก ลิฟวิงสตันจึงนิยาม

เสียใหม่ว่า



$$D_x^2 = \frac{\sum (x - c_x)^2}{N}$$

และ

$$D_{(x,y)} = \frac{\sum (x - c_x)(y - c_y)}{N}$$

เมื่อ c_x และ c_y เป็นคะแนนเกณฑ์มาตรฐานของแบบสอบอิงเกณฑ์ที่เป็นแบบสอบ
 คูนานา จากนั้นลิฟวิ่งส์ตันก็ใช้วิธีการของทฤษฎีคังเคิมในการพัฒนาสูตรจนได้คังสมการข้างต้น
 ซึ่งเมื่อพิจารณาจะพบว่า ถ้ายิ่ง $u - c$ แตกต่างกันมาก ค่าความเที่ยงที่ได้ตามสูตรคังกล่าว
 ก็จะเพิ่มตามไปด้วย และแม้ว่าค่าความเที่ยงในแนวทฤษฎีคังเคิมจะมีค่าเป็นศูนย์ แต่เมื่อใช้สูตร
 ของลิฟวิ่งส์ตัน ค่าความเที่ยงที่ได้จะยังคงมีค่ามากกว่าศูนย์ จากถรกที่มีค่า $(u - c)^2$ เพิ่ม
 เข้าไปอีกทั้งเศษและส่วน ทำให้ค่าความเที่ยงที่ได้มันแปรไปตามค่า $(u - c)^2$ ด้วย มีนัก-
 วิชาการบางท่าน เช่น เชฟเวลดสัน, บล็อก และ รวิทย์ (Shavelson, Block and
 Ravitch 1972) ไม่เห็นด้วยว่าวิธีการนี้จะเป็นการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์

ทอมมา เบรนนัน และ เคน (Brennan and Kane 1977) ได้เสนอแนวทางการ
 หาค่าความเที่ยงที่พัฒนาจากแนวคิดของทฤษฎีการสรุปให้เป็นทั่วไป (Generalizability
 Theory) เป็นสำคัญ โดยมีแนวคิดว่า แบบสอบอิงเกณฑ์นั้นข้อสอบที่สุ่มขึ้นมาใช้ทดสอบเป็นเพียง
 ส่วนย่อยของข้อสอบทั้งหมดที่ดีเป็นประชากรข้อสอบ และข้อสอบที่สุ่มขึ้นมาใช้นั้นย่อมจะมีความแตก
 ต่างกันและจะมีค่าคลาดเคลื่อนอันเนื่องมาจากการสุ่มข้อสอบด้วยส่วนหนึ่ง ซึ่งจะกลายมาเป็น
 แหล่งความแปรปรวนคลาดเคลื่อนที่จะต้องคำนึงถึงด้วย เบรนนัน เรียกค่าที่ได้จากวิธีการที่พัฒนา
 ขึ้นมานี้ว่า "Index of Dependability"

แนวคิดของเบรนนัน เริ่มต้นจากการอธิบายว่า คะแนนของผู้เข้าสอบคนที่ p ใน
 ข้อสอบที่ i จะเขียนเป็นสมการเส้นตรง (Linear Model) แยกส่วนประกอบของคะแนน
 ออกได้เป็นดังนี้

$$X_{pi} = U + \tau_p + \beta_i + \tau B_{pi} + e_o(pi)$$

- เมื่อ U คือค่าเฉลี่ยในประชากรผู้เข้าสอบและประชากรผู้สอบ
- τ_p คือผลที่เกิดจากผู้เข้าสอบคนที่ p
- β_i คือผลที่เกิดจากข้อสอบข้อที่ i
- τB_{pi} คือปฏิกริยาสัมพันธ์ระหว่างผู้เข้าสอบคนที่ p และข้อสอบข้อที่ i
- $e_o(pi)$ คือค่าความคลาดเคลื่อนที่เกิดขึ้น

ในการพัฒนาสูตรเบรนนันก็ยังคงยึดแนวคิดที่ว่า การคำนวณความแปรปรวนของคะแนนในแบบสอบอิงเกณฑ์ ควรจะคิดจากผลทางของค่าคะแนนที่สอบได้กับจุดเกณฑ์มาตรฐาน เช่นเดียวกันกับวิธีการของลิฟวิงสตัน ในตอนท้ายสุดของการพัฒนาสูตรจะได้สมการเพื่อหาค่าความเที่ยง คือ

$$M(C) = \frac{\sigma^2(\tau) + (U - C)^2}{\sigma^2(\tau) + (U - C)^2 + \frac{1}{n_i} [\sigma^2(\beta) + \sigma^2(\tau B, e)]}$$

(Brennan and Kane 1977 : 280)

- โดยที่ $M(C)$ คือค่า
- $\sigma^2(\tau)$ คือความแปรปรวนของคะแนนที่แท้จริง
- $\sigma^2(\beta)$ คือความแปรปรวนอันเนื่องมาจากข้อสอบ
- $\sigma^2(\tau B, e)$ คือความแปรปรวนอันเนื่องมาจากปฏิกริยาสัมพันธ์ระหว่างคนกับข้อสอบ รวมกับความแปรปรวนความคลาดเคลื่อน
- U คือค่ามัธยิมเลขคณิตของกลุ่ม
- C คือค่าเกณฑ์มาตรฐานที่กำหนดขึ้น
- n คือจำนวนข้อสอบทั้งหมด

ทั้งนี้การหาค่า $\sigma^2(\tau)$ $\sigma^2(\beta)$ และ $\sigma^2(\tau\beta, e)$ ได้มาจากการวิเคราะห์ความแปรปรวนค่าความเที่ยงจากสูตรของเบรนนันจะมีค่าต่ำกว่าหรือเท่ากับค่าความเที่ยงที่คำนวณได้จากสูตรของลิฟวิงสตัน ทั้งนี้เนื่องจากว่าเบรนนันยึดหลักของทฤษฎีการสรุปให้เป็นทั่วไปถือว่าข้อสอบในโมเดลการวิเคราะห์เป็นตัวอย่างสุ่มตัวหนึ่ง ค่าความแปรปรวนที่สังเกตได้ จึงแยกออกเป็น

$$\sigma_o^2 = \sigma^2(\tau) + [\sigma^2(\beta) + \sigma^2(\tau\beta, e)] / n$$

ซึ่งจะเห็นว่าค่า $\sigma^2(\beta)$ เป็นแหล่งความแปรปรวนความคลาดเคลื่อนที่เพิ่ม

3. แนวคิดของการประมาณค่าคะแนนที่แท้จริงของผู้สอบ

หลักการของแนวคิดนี้ไม่ใ้มุ่งหาค่าความเที่ยงออกมาเป็นตัวเลขโดยตรงเช่นวิธีการอื่น แต่มุ่งที่จะหาช่วงความเชื่อมั่นของคะแนนที่แท้จริงของผู้สอบเป็นสำคัญ ในประเด็นของการคำนวณจึงมุ่งเน้นการหาค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) โดยวิธีการคำนวณยังแยกออกได้เป็น 2 แนวทางหลัก คือ

1. การหาค่าความคลาดเคลื่อนมาตรฐานของผู้สอบแต่ละคน
2. การหาค่าความคลาดเคลื่อนมาตรฐานของกลุ่ม

การหาค่าความคลาดเคลื่อนมาตรฐานของผู้สอบแต่ละคนนั้น ใช้แนวคิดของการแจกแจงแบบทวินามเป็นหลักในการหา ซึ่งมีสูตรที่ใช้ในการคำนวณ คือ

$$\hat{\sigma}_p = \sqrt{\frac{pq}{n - 1}}$$

- เมื่อ p คือความคลาดเคลื่อนมาตรฐานในการวัด
- p คือสัดส่วนของคะแนนที่สอบได้
- q คือ 1 - p
- n คือจำนวนข้อสอบทั้งหมดในแบบสอบ

หรืออาจจะเขียนอยู่ในรูปของคะแนนเต็มที่ได้จากการสอบ คือ

$$S.E.(X_a) = \frac{X_a(n - X_a)}{n - 1}$$

เมื่อ X_a คือคะแนนที่ได้จากการสอบ

n คือจำนวนข้อสอบทั้งหมดในแบบสอบ

สำหรับการหาค่าความคลาดเคลื่อนมาตรฐานของกลุ่ม มีดัชนีที่ใช้อยู่ 2 ค่า คือ

σ_E^2

และ σ_A^2 โดยที่

σ_E^2 เป็นการประมาณค่าความคลาดเคลื่อนมาตรฐานในแนวคิดทฤษฎีคลังเก็บ

σ_A^2 เป็นการประมาณค่าความคลาดเคลื่อนมาตรฐานในแนวคิดของทฤษฎีการ

สรุปให้เป็นทั่วไป

ทั้งนี้ เบิร์ก (Berk 1980 a : 344) ได้สรุปไว้ว่า การที่จะเลือกใช้ค่าความคลาดเคลื่อนมาตรฐานแบบใดแบบหนึ่งนั้น ขึ้นอยู่กับนิยามความคลาดเคลื่อนในสภาพการณ์ของการตรวจสอบ เป็นหลักการสำคัญ

เทราบบ์ และ โรว์เลย์ (Traub and Rowley 1980 : 520 - 522) ได้กล่าวไว้ว่า การพัฒนาวิธีการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์ใหม่ส่วนทำให้ความหมายของคำว่า "ความเที่ยง" เปลี่ยนแปลงไป เช่น การหาค่าความเที่ยงในแนวคิดของการหาค่าความคงที่ในการจัดประเภทบุคคลนั้น ไม่มีส่วนเกี่ยวข้องกับคะแนนที่แท้จริงของผู้สอบเลย การหาค่าความเที่ยงของแนวคิดการหาความแตกต่างของคะแนนที่สอบได้กับคะแนนเกณฑ์มาตรฐานก็ไม่ได้สื่อความหมายถึงวิธีการประมาณค่าคะแนนที่แท้จริง ดังนั้น เพื่อให้เกิดความชัดเจนในการสื่อความหมาย ผู้ที่กล่าวถึงค่าความเที่ยงจึงควรระบุให้ชัดว่า ท้องการสื่อความหมายถึงค่าความเที่ยงประเภทใด หรือใช้สูตรใดในการคำนวณ นอกจากนี้ แท้จริงแล้ว ถ้าแบบสอบที่สร้างขึ้นมานั้นมุ่งวัดตัวแปรที่มีลักษณะต่อเนื่อง (Continuum) และผลการสอบก็เป็นไปเพื่อหาตำแหน่งที่ความสามารถที่แท้จริงของบุคคลในช่วงของความต่อเนื่องดังกล่าวแล้ว การหาค่าความเที่ยงในแนวคิดทฤษฎีคลังเก็บก็สามารถนำมาใช้ได้ และยังสื่อความหมายได้ถึงคะแนนที่แท้จริงของบุคคลอีกด้วย

1.6 ปัญหาการกำหนดความยาวของแบบสอบถามเชิงคุณภาพ

ดังได้กล่าวไว้แล้วในตอนต้นของบทนำว่า การกำหนดความยาวของแบบสอบถามเชิงคุณภาพให้มีขนาดพอเหมาะ คือ ประกอบขึ้นด้วยจำนวนข้อสอบไม่มากนัก แต่มีความสามารถจับประเภทบุคคลตามระดับการรอบรู้ได้อย่างมีความถูกต้องสูง ภายใต้ข้อควรระวังความคลาดเคลื่อนและเกณฑ์มาตรฐานที่กำหนดขึ้นไว้แล้ว ยังไม่มีวิธีการที่เด่นชัด เมื่อศึกษาจากงานที่เกี่ยวข้องพบว่า มาตราการกำหนดความยาวของแบบสอบถามที่มีการศึกษาและนำเสนอมาแล้วมีอยู่ 4 วิธีด้วยกัน คือ

- ก. วิธีการของมิลล์แมน
- ข. วิธีการวิลคอค
- ค. วิธีการของเบเชียนโมเดล
- ง. วิธีการใช้ทฤษฎีการตอบสนองข้อสอบ

แต่ละวิธีมีลักษณะแนวคิดที่สำคัญดังต่อไปนี้

- ก. วิธีการของมิลล์แมน

มิลล์แมนได้ใช้หลักการแจกแจงแบบทวินามมาเป็นหลักในการกำหนดความยาวของแบบสอบถาม โดยให้นำเสนอในลักษณะของตารางสำเร็จรูป ดังเช่น

ตารางที่ 6 การกำหนดความยาวของแบบสอบถามเชิงคุณภาพโดยวิธีการของมิลล์แมน

จุดตัด	จำนวนข้อสอบ	ระดับความสามารถที่แท้จริง								
7 ใน 10	10	40	50	60	65	75	80	90		
		5	17	38	51	22	12	1		

(Millman in Popham 1974 : 342)

จากตาราง ตัวเลขในตารางเป็นค่าของโอกาสความคลาดเคลื่อนในการจับประเภทบุคคลตามระดับการรอบรู้ที่อาจจะเกิดขึ้นได้ โดยมีความหมายดังนี้

- ค่าที่อยู่ทางขวามือของเส้นประ หมายถึง โอกาสที่จะจับบุคคลที่มีความสามารถที่แท้จริงสูงกว่าเกณฑ์มาตรฐาน เป็นผู้ไม่รอบรู้ ดังเช่นจากตาราง จะมีความหมายว่า ถ้าออกข้อสอบ 10 ข้อ กำหนดเกณฑ์มาตรฐานเป็น 0.70 ผู้สอบที่มีความสามารถที่แท้จริงเป็น 0.75 0.80 และ 0.90 จะมีโอกาสได้รับการจัดประเภทเป็นผู้ไม่รอบรู้ อยู่ 0.22 0.12 และ 0.01 ตามลำดับ

- ค่าที่อยู่ทางซ้ายมือของเส้นประ หมายถึง โอกาสที่จะจับบุคคลที่มีความสามารถที่แท้จริงต่ำกว่าเกณฑ์มาตรฐาน เป็นผู้รอบรู้ ดังเช่นจากตาราง ผู้สอบที่มีความสามารถที่แท้จริงเป็น 0.40 0.50 0.60 และ 0.65 จะมีโอกาสได้รับการจัดประเภทเป็นผู้รอบรู้ อยู่ 0.05 0.17 0.38 และ 0.51 ตามลำดับ

สำหรับแนวคิดในการคำนวณนั้น มีปรากฏในคำอธิบายของแฮมเบิลตัน (Hambleton in Berk 1984 : 147) ซึ่งจากคำอธิบายดังกล่าว สามารถที่จะแสดงการคำนวณให้ดูได้ถึงที่มาของค่าในตาราง ดังตัวอย่าง คือ

ผู้สอบที่มีความสามารถที่แท้จริงเป็น 0.50 จะมีโอกาสตอบข้อสอบถูกต้องตั้งแต่ 7 ข้อขึ้นไป ในจำนวน 10 ข้อ จะคำนวณได้จากวิธีการดังต่อไปนี้

$$\begin{aligned} \text{โอกาสตอบถูก 7 ใน 10} &= \begin{bmatrix} 10 \\ 7 \end{bmatrix} \frac{5^7}{10^7} \frac{5^3}{10^3} + \begin{bmatrix} 10 \\ 8 \end{bmatrix} \frac{5^8}{10^8} \frac{5^2}{10^2} \\ \text{ข้อขึ้นไป ของผู้สอบที่มี} & \\ \text{ความสามารถตรง 0.50} &+ \begin{bmatrix} 10 \\ 9 \end{bmatrix} \frac{5^9}{10^9} \frac{5^1}{10^1} + \begin{bmatrix} 10 \\ 10 \end{bmatrix} \frac{5^{10}}{10^{10}} \frac{5^0}{10^0} \\ &= 0.17 \end{aligned}$$

อย่างไรก็ดี แนวคิดของมิลล์แมน มีข้อที่นำพิจารณาอยู่ 2 ประการ คือ

1. ในการทดสอบตามปกติทั่วไป จะไม่สามารถทราบค่าระดับความสามารถที่แท้จริงของผู้สอบได้ แม้มิลล์แมนใช้วิธีการคำนวณโดยยึดหลักการทราบค่าความสามารถที่แท้จริงของผู้สอบ ซึ่งโนวิกและเลวิส (Novick and Lewis in Harris 1974 : 142) เห็นว่าวิธีการนี้ไม่ถูกต้อง

2. ตามวิธีการของมิลล์แมน ถ้าต้องการให้เกิดความคลาดเคลื่อนในการจัดประเภทบุคคลอื่นเนื่องมาจากโอกาสแต่เพียงอย่างเดียว จะต้องใช้แบบสอบที่มีความยาวมาก ในขณะที่

ที่ Novick และ Lewis กล่าวว่า ความยาวของแบบสอบถามเชิงทัศนคติที่สั้นที่สุดไม่ควรจะเกิน 20 ข้อ (Novick and Lewis 1974 : 31)

ข. วิธีการของวิลคอก

วิลคอกได้พัฒนาวิธีการกำหนดความยาวของแบบสอบถามเชิงทัศนคติ โดยกล่าวอ้างถึงงานเขียนของเฟเนออร์ (Wilcox 1976) ที่เสนอแนวคิดของการกำหนดความยาวของแบบสอบถามเชิงทัศนคติ โดยใช้การแจกแจงแบบทวินามเช่นกัน วิธีการนี้ เริ่มจากค่าเกณฑ์มาตรฐานที่กำหนดขึ้น ($\%_0$) จะมีการประมาณช่วงออกไปเป็น $\%_0 - c$ และ $\%_0 + c$ โดยถือว่า ถ้าผู้สอบคนใดมีความสามารถที่แท้จริงอยู่ในช่วง $\%_0 \pm c$ เมื่อมีการจับประเภทบุคคลที่มีความสามารถในช่วงนี้ผิด จะไม่ถือเป็นความผิดพลาด หรือเรียกช่วงดังกล่าวว่า "Indifference Zone" ทั้งนี้แล้ว โอกาสที่ทัศนคติจะถูกในการให้ผู้สอบสอบตกจะมีค่า

$$\alpha = \sum_{x=0}^{n_0-1} \binom{n}{x} (\%_0 + c)^x (1 - \%_0 + c)^{n-x}$$

และโอกาสที่จะทัศนคติจะถูกในการให้ผู้สอบสอบผ่านจะมีค่า

$$\beta = \sum_{x=n_0}^n \binom{n}{x} (\%_0 + c)^x (1 - \%_0 - c)^{n-x}$$

เมื่อ x เป็นจำนวนข้อสอบที่ถือเป็นจุดเกณฑ์มาตรฐาน

n เป็นจำนวนข้อสอบทั้งหมด

n_0 เป็นจำนวนข้อสอบที่ผู้สอบตอบถูก

ทั้งนี้ การกำหนดจำนวนข้อสอบที่เหมาะสมของแบบสอบ จะกระทำตรงจุด $\alpha \gg P_1^*$ และ $\beta \gg P_2^*$ เมื่อ P_1^* และ P_2^* เป็นโอกาสในการที่จะจับประเภทบุคคลถูก โดยที่ค่า P_1^* และ P_2^* จะต้องมีค่าอยู่ในช่วง 0.50 - 1.00

อย่างไรก็ดี วิธีการนี้ยังคงเป็นการคิดคำนวณทางคณิตศาสตร์ โดยไม่มีข้อมูลเชิงประจักษ์ยืนยัน โดยวิธีการคำนวณทางคณิตศาสตร์ดังกล่าวก็ไม่สามารถจะบอกได้ว่า โอกาสที่ผู้สอบจะตกอยู่ในเขตที่ถือว่าไม่เกิดความผิดพลาดในการทัศนคติ (Indifference Zone) มีอยู่เท่าไร ทั้ง

วิธีการทางคณิตศาสตร์ดังกล่าวก็ไม่ใ้มีการกล่าวหาเพียง ไปถึงลักษณะของ โดเมนที่มุ่งสร้างแบบสอบถาม ถ้าโดเมนมีขนาดต่างกัน จะมีผลต่อการกำหนดความยาวที่เหมาะสมของแบบสอบถามอย่างไร

ค. วิธีการของเบเซียนของโมเดล

มิลล์แมนได้อ้างถึงงานเขียนของนอวิกและเลวิส (Millman in Popham : 354) ที่ได้ประยุกต์หลักการทางสถิติที่เรียกว่า เบเซียนโมเดลมากำหนดความยาวของแบบสอบถามเชิงเกณฑ์ เช่นจากตาราง

ตารางที่ 7 การกำหนดความยาวของแบบสอบถามเชิงเกณฑ์โดยวิธีการของเบเซียนโมเดล

		อัตราการสูญเสีย			
\mathcal{T}_0	$\Sigma(\mathcal{T}^c)$	1.5	2	2.5	3
70%	70%	6/8(75%)	10/13(77%)	11/14(79%)	12/15(80%)

เมื่อกำหนด \mathcal{T}_0 คือความสามารถ หรือ ระดับเกณฑ์มาตรฐาน
 $(\mathcal{T})^c$ คือค่าคะแนนเฉลี่ยของกลุ่มที่ใช้เป็นข้อมูลในตอนต้น (Prior Distribution) ตามหลักการของเบเซียนโมเดล

อัตราการสูญเสีย คือ สัดส่วนค่าน้ำหนักระหว่าง การให้ผู้ไม่รอบรู้สอบผ่าน ต่อ การให้ผู้รอบรู้สอบตก

จากตารางดังกล่าว จะมีความหมายว่า ถ้ากำหนดเกณฑ์มาตรฐานการสอบผ่านเป็น 0.70 การประมาณค่าคะแนนของกลุ่มได้ค่าเป็น 0.70 ถ้ากำหนดอัตราการสูญเสียเป็น 1.5 จะต้องใช้แบบสอบถามจำนวน 8 ข้อ และผู้สอบจะต้องทำได้ 6 ข้อขึ้นไป จึงจะถือว่าเป็นผู้รอบรู้

แต่ทั้งนี้ เงื่อนไขของการคำนวณตามโมเดลนี้มีอยู่ว่า จะคงสามารถระบุลักษณะการแจกแจงของข้อมูลในทอนตันได้ ซึ่งประเด็นนี้ วิลคอก (Wilcox 1980 : 430) ได้กล่าวไว้ว่า มีผู้โต้แย้งอยู่มาก เช่น บาร์เรตต์ (Barette) กล่าวไว้ว่า หลักการของวิชาสถิติคือสามารถทดสอบได้ แต่ข้อมูลที่ใช่เป็นส่วนของการแจกแจงในทอนตัน เป็นสิ่งที่ได้มาจากการกำหนดขึ้นเท่านั้น ทั้งไม่สามารถทดสอบได้ทางหลักการของวิชาสถิติ

ง. การใช้ทฤษฎีทดสอบสองข้อสอบ

แฮมเบิลตัน, มิลล์ และ ซิมอน (Hambleton, Mills and Simon 1983) ได้ศึกษาข้อมูลที่สร้างขึ้นจากเครื่องคอมพิวเตอร์และใช้ทฤษฎีการทดสอบสองข้อสอบศึกษาหาค่าตอบในประเด็นต่างๆต่อไปนี้

1. ผลกระทบของการใช้ข้อสอบที่มีค่าความยากต่างกันต่อการกำหนดความยาวของแบบสอบ
2. วิธีการต่างๆในการสุ่มข้อสอบขึ้นมาทดสอบ ที่มีผลต่อความเที่ยงที่ใช้ในความหมายของความคงที่ในการจัดประเภทบุคคล (คัทนี่แคปป์)
3. ผลกระทบของค่าอำนาจจำแนกที่มีต่อความยาวของแบบสอบอิง เกณฑ์จากการศึกษาพบว่า
 1. แบบสอบที่มีความยาวมากขึ้น จะมีค่าความเที่ยงเพิ่มขึ้น
 2. แบบสอบที่มีค่าความยากใกล้เคียงกัน จะมีค่าความเที่ยงเพิ่มมากขึ้น ๓ ที่ระดับความยาวเท่ากัน
 3. แบบสอบที่มีค่าอำนาจจำแนกสูงขึ้น จะมีค่าความเที่ยงสูงกว่าแบบสอบที่เป็นคู่ขนานแบบสุ่มที่มีค่าการกระจายของค่าความยากเท่ากัน และมีความยาวเท่ากัน

แฮมเบิลตันและกรูเจอร์ (Hambleton and Gruijter 1983) ได้ศึกษามาตรการการกำหนดความยาวของแบบสอบอิง เกณฑ์ที่สร้างขึ้นจากข้อสอบที่มีการกระจายของค่าความยากและค่าอำนาจจำแนกต่างกัน โดยใช้ทฤษฎีการทดสอบสองข้อสอบ โดยแนวคิดนี้จะมีความสัมพันธ์กับการนับคะแนนในแบบสอบที่ใช้กันมาก่อนว่า

$$\pi = \frac{1}{m} \sum_{i=1}^m P_i(\theta)$$

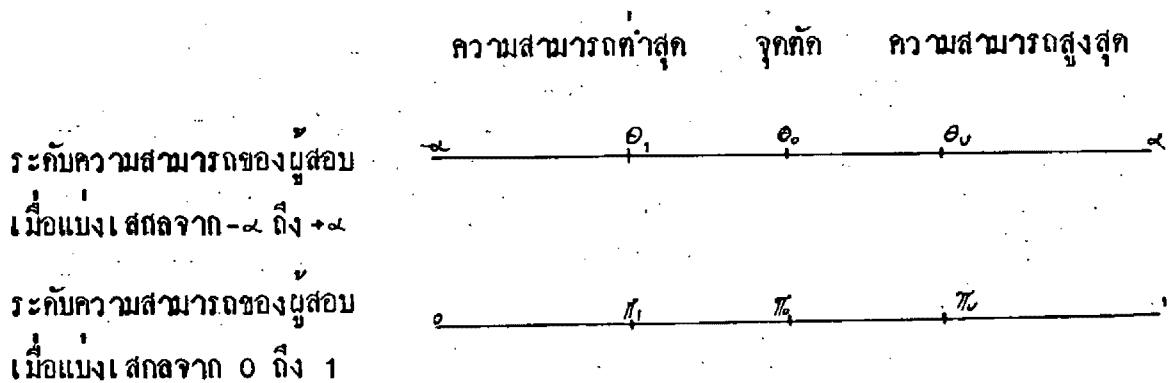
เมื่อ P_i เป็นโอกาสที่ผู้สอบซึ่งมีความสามารถตรงกับ θ จะทำข้อสอบข้อหนึ่งๆ ได้ตามทฤษฎีผลรวมของโอกาสที่ผู้สอบจะทำข้อสอบแต่ละข้อถูก หากด้วยจำนวนข้อสอบทั้งหมด จะมีค่าเท่ากับสัดส่วนของข้อสอบที่บุคคลนั้นจะทำข้อสอบถูกในโคเมนที่กำหนด และระดับของเกณฑ์มาตรฐานก็มีความสัมพันธ์กันในลักษณะเดียวกัน คือ

$$\pi_0 = \frac{1}{m} \sum_{i=1}^m P_i(\theta_0)$$

เมื่อ P_i เป็นโอกาสที่ผู้สอบซึ่งมีความสามารถตรงกับเกณฑ์มาตรฐาน หรือจะทำข้อสอบข้อหนึ่งๆ ได้

จากนั้นก็ใช้หลักการว่า ถ้าหากสามารถกำหนดให้การกระจายของระดับความสามารถของบุคคลในเรื่องที่สนใจจะทดสอบเป็นจุดต่อเนื่องในแนวเส้นตรง จากกรณีที่ไม่มีความรู้เลยไปจนถึงมีความรู้ในเรื่องนั้นเป็นอย่างดี จะมีลักษณะดังแผนภูมิต่อไปนี้

แผนภูมิที่ 6 การกระจายความสามารถของผู้สอบในระดับต่างๆ ณ จุดต่อเนื่องในแนวเส้นตรงตามทฤษฎีการตอบสนองข้อสอบ



บริเวณให้ตก Indifference บริเวณให้ผ่าน
Zone

(ดู Hambleton in Berk 1984 : 155)

จากแนวคิดดังกล่าว เมื่อกำหนดให้

P_1 เป็นโอกาสของความคลาดเคลื่อนที่เกิดจากให้ผู้รอบรู้สอบตก

P_u เป็นโอกาสความคลาดเคลื่อนที่เกิดจากการให้ผู้ไม่รอบรู้สอบผ่าน

P^* เป็นโอกาสความคลาดเคลื่อนที่ถือเป็นเกณฑ์ หรือ คือโอกาสความคลาดเคลื่อนสูงสุดที่ยอมให้เกิดขึ้นได้

จากหลักการดังกล่าว วิธีการที่แฮมเบิลตันและกรูจเตอร์ใช้ในการกำหนดความยาว

จะกระทำดังนี้

$$P_m = \text{Max}(P_1, P_u) \leq P^*$$

ซึ่งผลการศึกษาโดยไซมอนจากเครื่องคอมพิวเตอร์ พบว่า

1. การให้ข้อสอบที่มีระดับความยากพอเหมาะกับระดับความสามารถของผู้สอบ จะลดความคลาดเคลื่อนลงไ้มากกว่าการให้แบบสอบที่สร้างขึ้นจากการสุ่มข้อสอบขึ้นมาให้ทดสอบ
2. แบบสอบที่มีความยาวเท่ากัน ข้อสอบที่มีค่าความยากเท่ากันและตรงกับระดับความสามารถของผู้สอบ แบบสอบที่ประกอบด้วยข้อสอบที่มีค่าอำนาจจำแนกสูงกว่า จะลดความคลาดเคลื่อนต่ำกว่า

ในปี 1983 ฮาลาดิโนนา และ รอยด์ (Haladyna and Roid 1983) ได้ศึกษาเปรียบเทียบผลของความคลาดเคลื่อนในการประมาณค่าความสามารถของบุคคลจากการใช้แบบสอบต่างกัน 2 แบบ คือ แบบแรกเป็นแบบสอบที่สุ่มข้อสอบขึ้นมาจากระชากรข้อสอบที่มีอยู่ทั้งหมดกับแบบหลัง เป็นการเลือกข้อสอบให้เหมาะกับระดับความสามารถของผู้สอบ โดยศึกษาจากรายวิชาต่างๆ 4 รายวิชา ซึ่งประกอบด้วยข้อสอบที่ผ่านการวิเคราะห์ด้วยวิธีไรโมเดลมาแล้ว ข้อสอบเหล่านี้เป็นประชากรข้อสอบทั้งหมดในโคเมน จากนั้นจึงได้ทดลองสุ่มสร้างแบบสอบที่มีจำนวนข้อสอบเป็น 10, 20, 30 และ 40 ข้อตามลำดับ แล้วศึกษาเปรียบเทียบอัตราความแตกต่างเฉลี่ยระหว่างคะแนนที่สอบได้กับคะแนนโคเมนที่ประมาณได้ (The Average Absolute Difference between Observed and Domain Scores) ที่ย่อ AAD จากแบบสอบที่มีความยาวระดับต่างๆที่กำหนด หาค่าส่วนเบี่ยงเบนมาตรฐาน (เรียกค่าที่ได้อีกว่า อัตราส่วน

AAD/SD) โดยในการสร้างแบบสอบถาม จะม่วิธีการสร้าง 4 วิธี คือ

1. สุ่มข้อสอบขึ้นมาจากโคเมน
2. เลือกข้อสอบใหม่มีความยากตรงกับระดับความสามารถของผู้สอบ
3. เลือกข้อสอบใหม่ค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบ

เลือกน้อย

4. เลือกข้อสอบใหม่ค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบ

จากการศึกษาพบว่า การสอบที่จัดให้ข้อสอบมีค่าความยากตรงกับระดับความสามารถของผู้สอบ จะทำให้ได้ค่าอัตราส่วน AAD/SD ที่ต่ำสุด แบบสอบที่ได้รับการสุ่มข้อสอบขึ้นมา มีค่าที่ขึ้นถึงกล่าวว่าเป็นลำดับที่สอง และแบบสอบที่มีค่าความยากกระจายมาก จะยังมีค่าที่ขึ้นถึงกล่าวสูงมากยิ่งขึ้น

ผู้เขียนทั้งสองได้อภิปรายไว้ว่า ในการดำเนินการสอบ ไม่ว่าจะใช้วิธีการใดจัดระดับความยากให้เหมาะกับระดับความสามารถของผู้สอบ หรือการสุ่มข้อสอบ ขึ้นมาสอบต่างก็ต้องการข้อสอบที่วัดในเนื้อหาเดียวกัน การสอบแบบจัดระดับความยากของข้อสอบให้เหมาะกับระดับความสามารถของผู้สอบจะใช้ได้ดีในกรณีที่ต้องการวัดคุณลักษณะความสามารถที่ต่อเนื่องกัน การทดสอบจะเป็นการพยายามหาค่าแทนที่ของระดับความสามารถของบุคคลในช่วงความต่อเนื่องนั้น แต่อย่างไรก็ดี เนื่องจากวิธีการสุ่มข้อสอบก็ยังคงให้ผลดีเป็นลำดับที่สอง และมีความสะดวกในการปฏิบัติ จึงน่าจะใช้เทคนิคทั้งสองในการสร้างแบบสอบอิงเกณฑ์ การที่จะตัดสินว่าวิธีการใดจะดีกว่า จะต้องรอผลจากประโยชน์ที่ได้จากการทดสอบจริง

โดยสรุปแล้ว การนำเอาทฤษฎีการทดสอบข้อสอบ มาใช้กับแนวคิดของการกำหนดเกณฑ์มาตรฐานของแบบสอบอิงเกณฑ์ได้ข้อค้นพบที่สำคัญว่า การคัดเลือกข้อสอบโดยให้ผู้สอบได้ทำข้อสอบที่มีความยากตรงกับความสามารถของเขา ค่าความคลาดเคลื่อนในการประมาณค่าความสามารถจะต่ำกว่าการสุ่มข้อสอบขึ้นมาใช้เพื่อการทดสอบ

แต่อย่างไรก็ดี การประยุกต์หลักการดังกล่าว มีข้อจำกัด คือ

1. แบบสอบอิงกลุ่มมุ่งเน้นที่การกำหนดให้ข้อสอบแต่ละข้อวัดได้ตรงตามวัตถุประสงค์ที่กำหนดเป็นสิ่งสำคัญอันดับแรก ส่วนคุณสมบัติด้านค่าประจำข้อสอบ เป็นเพียงสิ่งช่วยเสริมเท่านั้น แต่หลักการของโมเดลนี้ชี้เอาค่าประจำข้อสอบเป็นจุดหลักของการวิเคราะห์ ซึ่งไม่ใช่หลักการ

สำคัญของแบบสอบอิงเกณฑ์ ฮาลาดิโนนา และ รอยด์ (Haladyna and Roid 1981 : 51) ก็คือกล่าวถึงความสำคัญของหลักการนี้

2. ลักษณะการใช้ทฤษฎีการตอบสนองข้อสอบนั้น เหมาะสมกับการสอบเพื่อบ่งชี้ระดับความสามารถของผู้สอบในเรื่องใด เรื่องหนึ่งซึ่งมีลักษณะเป็นการต่อเนื่องของความสามารถต่ำสุดไปหาความสามารถสูงสุด การทดสอบจะเป็นไปเพื่อระบุตำแหน่งที่ตั้งของความสามารถของผู้สอบในช่วงความต่อเนื่องนั้น เหตุนี้แบบสอบอิงเกณฑ์จะใช้กับโมเดลจะต้องเป็นแบบสอบอิงเกณฑ์ที่มีลักษณะเฉพาะที่เหมาะสมกับโมเดล ซึ่งแฮมเบิลตัน (Hambleton in Berk 1984 : 164) ก็โต้กล่าวไว้ว่า โมเดลนี้ใช้ได้กับแบบสอบในบางกรณีเท่านั้น นอกจากนั้น ในกรณีที่การตรวจสอบพบว่าข้อสอบในแบบสอบไม่สามารถจะใช้ได้กับโมเดลการวิเคราะห์นั้นแล้ว ก็ควรจะคงข้อสอบไว้ และตัดโมเดลที่ใช้ในการวิเคราะห์ออกไป (Wilcox 1980 : 443)

แฮมเบิลตัน (Hambleton in Berk 1984 : 144 - 168) ได้สรุปรวบรวมแนวคิดต่างๆทั้งหมด เกี่ยวกับการกำหนดความยาวของแบบสอบอิงเกณฑ์ที่กระจุกกระจายเข้าไว้ด้วยกัน แต่อย่างไรก็ดี ในตอนท้ายของบทความ แฮมเบิลตันก็โต้กล่าวไว้ว่า ในทางปฏิบัตินั้นแทบจะไม่มี การนำเอาแนวคิดใดๆ เกี่ยวกับการกำหนดความยาวไปใช้จริง วิธีการกำหนดความยาวของแบบสอบอิงเกณฑ์ในทางปฏิบัติจะกำหนดจากข้อจำกัดต่างๆในการทำงาน เช่น เวลาที่มีอยู่สำหรับการจัดสอบ เป็นต้น ทั้งที่จริงแล้วการกำหนดความยาวควรจะได้นิ่งถึงปัจจัยอื่นอีกด้วย

1.7 พัฒนาการของแบบสอบอิงเกณฑ์ในประเทศไทย

การศึกษาวิจัยเกี่ยวกับแบบสอบอิงเกณฑ์ในประเทศไทย นับได้ว่า เพิ่งอยู่ในระยะเริ่มต้นเท่านั้น งานที่ค้นคว้าได้ทั้งหมดจะปรากฏอยู่ในรูปของวิทยานิพนธ์ในระดับบัณฑิตศึกษา ลักษณะของการศึกษาค้นคว้าที่สำคัญมีอยู่ 2 ลักษณะ คือ ลักษณะแรกจะเน้นหนักไปในทางการค้นคว้าเพื่อความก้าวหน้าทางวิชาการ อีกลักษณะหนึ่งจะมุ่งเน้นไปในทางการสร้างแบบสอบอิงเกณฑ์เพื่อใช้ประกอบการเรียนการสอนในระดับชั้นต่างๆ

งานค้นคว้าเกี่ยวกับการสร้างแบบสอบอิงเกณฑ์ชิ้นแรกเริ่มในปี 2522 กาญจนา วัชรสุนทร (2522) ได้ทำวิทยานิพนธ์ เรื่อง "การสร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์" ลักษณะของวิทยานิพนธ์เป็นการสร้างแบบสอบอิงเกณฑ์ในวิชาคณิตศาสตร์ เรื่อง สมการ ซึ่งใช้ในการเรียนการสอนสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยได้สร้างข้อสอบขึ้นตามรูปแบบเฉพาะประจำข้อสอบ โดยแบ่งแบบสอบออกเป็น 4 ส่วนย่อย คือ ความรู้เบื้องต้น เรื่องสมการและ

การแทนค่าสมการ ความรู้ในการแก้ปัญหาคณิตศาสตร์โดยใช้คุณสมบัติการบวก ลบ คูณ และหาร ความรู้ในเรื่องการเปลี่ยนโจทย์สมการเป็นสมการ และความรู้เกี่ยวกับการแก้โจทย์สมการ กลุ่มตัวอย่างที่ใช้ในการศึกษา เป็นนักเรียนจำนวน 156 คน จากโรงเรียนสันติราษฎร์ ในปีการศึกษา 2521 การวิเคราะห์ค่าประจำข้อสอบในการหาค่าความยาก และค่านีเอส (คือค่าความไวในการสอน) หาค่าความตรงโดยวิธีการของคาร์เวออร์ หาค่าความเที่ยงโดยวิธีการของ - ลิฟวิงสตัน ผลการทดสอบความแตกต่างของค่าที่ของผลสัมฤทธิ์ทางการเรียนในแบบสอบย่อย ทั้ง 4 ชุด ระหว่างหลังการเรียนกับก่อนการเรียนได้ค่าที่เป็น 47.79 36.73 36.94 และ 40.41 ตามลำดับ ซึ่งค่าที่ทดสอบทั้งหมดมีนัยสำคัญทางสถิติที่ระดับ 0.001

จากผลงานการสร้างแบบสอบอิงเกณฑ์ของ กาญจนา วัชรสุนทร ได้มีผู้นำไปศึกษาต่อในประเด็นอื่นๆ เพื่อความก้าวหน้าทางวิชาการอีกสองคน คือ ในปี 2523 ชมพู่ จันทอมรพร (2523) ได้ทำวิทยานิพนธ์เรื่อง "การใช้กระบวนการเชิงทฤษฎีการตัดสินใจของเบส์ในการกำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์" โดยได้ศึกษากับนักเรียนโรงเรียนบางปลาหมอ "สูงสุมาวณคุณวิทย์" จังหวัดสุพรรณบุรี จำนวน 80 คน ค่าในการหาคะแนนเกณฑ์มาตรฐานตามวิธีการของเบส์ ปรากฏว่าแบบสอบย่อยชุดต่างๆมีค่าคะแนนจุดตัดเป็น 0.60 0.50 0.50 และ 0.50 ตามลำดับ ในปี 2524 ประภา แก่นเพิ่ม (2524) ได้ทำวิทยานิพนธ์เรื่อง "ความถูกต้องในการกำหนดจุดตัดด้วยวิธีการกำหนดเกณฑ์ระดับผ่านต่ำสุด" โดยได้นำเอาแบบสอบที่ กาญจนา วัชรสุนทร สร้างขึ้น และ ชมพู่ จันทอมรพร ได้นำไปปรับแก้เล็กน้อย ไปศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2524 โรงเรียนหอวัง ที่มีผลการเรียนคาบเส้นตามทีครุระบุรีจำนวน 44 คน และนักเรียนที่มีผลการเรียนคาบเส้นในอิตาลีจำนวน 32 คน และศึกษากับนักเรียนโรงเรียนโยธินบูรณะ ที่มีผลการเรียนคาบเส้นตามทีครุระบุรีจำนวน 38 คน และนักเรียนที่มีผลการเรียนคาบเส้นตามผลการเรียนในอิตาลีอีกจำนวน 35 คน ลักษณะของการศึกษา เป็นการศึกษาเปรียบเทียบระหว่างการกำหนดเกณฑ์มาตรฐานขั้นต่ำจากการพิจารณาของผู้สอนกับผลการเรียนของนักเรียนที่ใช้เป็นตัวอย่างสอบ และลักษณะของการกำหนดเกณฑ์มาตรฐานขั้นต่ำจะกำหนดเพียงค่าเดียวจากคะแนนจากแบบสอบที่ฉบับมีจำนวน 40 ข้อ วิธีการกำหนดจุดตัดขั้นต่ำใช้ 2 วิธี คือ วิธีแรกจะหาค่า MPL_1 (Minimum Passing Level) ซึ่งเป็นการกำหนดจุดตัดขั้นต่ำตามแนวคิดของ กิลเบอรัท อีกวิธีหนึ่งจะเป็นการหา MPL_2 ซึ่งเป็นการกำหนดจุดตัดขั้นต่ำของมหาวิทยาลัยอิลลินอยส์ จากผลการศึกษา ปรากฏดังนี้

ตารางที่ 8 ผลการกำหนดจุดตัดขั้นต่ำของแบบสอบอิง เกณฑ์วิชาคณิตศาสตร์ เรื่องสมการ จากการศึกษาของ ประภา แก่นเพิ่ม

วิธีการกำหนดเกณฑ์มาตรฐานขั้นต่ำ	ค่าคะแนนที่ได้
MPL ₁	15.07
MPL ₂	12.61
การตรวจสอบผลการเรียนคาบเส้นตามที่ครูระบุ	
โรงเรียนหอวัง	15.89
โรงเรียนโยธินบูรณะ	14.58
การตรวจสอบผลการเรียนคาบเส้นตามผลการเรียนในอดีต	
โรงเรียนหอวัง	16.63
โรงเรียนโยธินบูรณะ	15.00

ซึ่งผู้วิจัยพบว่า ผลการเปรียบเทียบคะแนนจุดตัดขั้นต่ำจากวิธีการต่างๆ มีค่าแตกต่างกัน น้อยมาก กล่าวคือค่าที่ต่างกันมากที่สุดมีค่าไม่เกิน 10%

สำหรับการศึกษาในลักษณะของการสร้างแบบสอบอิง เกณฑ์นั้น สามารถเปรียบเทียบความปีการศึกษาที่ค่าเน้นการได้ดังต่อไปนี้

ปี 2524 สมถวิล วิจิตรวรรณ (2524) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิง เกณฑ์วิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 2" โดยศึกษากับนักเรียนชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2523 ในจังหวัดสมุทรปราการ จำนวน 370 คน ได้แบ่งแบบสอบออกเป็น 4 ฉบับย่อย คือการนับผล สัญญลักษณ์ การหาร โจทย์ปัญหาสมการ การหาค่าความตรงของแบบสอบใช้วิธีการของคาร์เวอ์ หาค่าความเที่ยงโดยการใช้สูตรของสับโคเวียค และในปีเดียวกันนี้ ไพฑูรย์ เวทการ (2524) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบสอบอิง เกณฑ์วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1" โดยศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โรงเรียนจอมสุรางค์อุปถัมภ์ จังหวัดพระนครศรีอยุธยา จำนวน 268 คน แบ่งแบบสอบออกเป็น 5 ฉบับย่อย คือ ความรู้พื้นฐานเรื่องอัตราส่วน อัตราส่วนอย่างต่ำ อัตราส่วนที่เท่ากัน อัตราส่วนกับการวัด และแบบทดสอบอัตรา การหาค่าความตรงใช้วิธีการของคาร์เวอ์ และหาค่าความเที่ยงโดยวิธีการของฮวน

ปี 2525 บุญเลิศ คำหอม (2525) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงเกณฑ์เรื่องสมการและอสมการ ชั้นมัธยมศึกษาปีที่ 2" โดยศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ของโรงเรียนวัดสามาซุกรักณาโคการาม จังหวัดสุพรรณบุรี จำนวน 240 คน แบ่งแบบสอบออกเป็น 4 ฉบับย่อย คือ การเปลี่ยนประโยคภาษาและประโยคสัญลักษณ์ คำตอบและกราฟของสมการและอสมการ การแก้สมการและอสมการ การแก้โจทย์ปัญหาสมการและอสมการ การหาค่าความตรงของแบบสอบใช้วิธีการของคาร์เวอว์ หาค่าความเที่ยงโดยวิธีการของสวามินาราน

ปี 2526 วุฒิคุณ เสาวลักษณ์ (2526) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์เรื่อง "ภาคตัดกรวย ชั้นมัธยมศึกษาปีที่ 4" โดยศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ปีการศึกษา 2525 ในจังหวัดประจวบคีรีขันธ์ จำนวน 137 คน แบ่งแบบสอบออกเป็น 11 ฉบับย่อย คือ การหาค่าความสัมพันธ์หรือสมการของกราฟวงกลม การหาโคออดิเนตของจุดศูนย์กลาง การหาจุดตัดของกราฟวงกลมกับกราฟวงกลม พาราโบลาที่มีจุดยอดที่จุด (0.0) พาราโบลาที่มีจุดยอดที่จุด (h, k) วงรีที่มีจุดศูนย์กลางที่จุด (0.0) วงรีที่มีจุดศูนย์กลางที่จุด (h, k) ไฮเพอร์โบลาที่มีจุดศูนย์กลางที่จุด (0.0) ไฮเพอร์โบลาที่มีจุดศูนย์กลางอยู่ที่จุด (h, k) ไฮเพอร์โบลามุมฉาก และการพิจารณาชนิดของกราฟจากความสัมพันธ์ โดยในการพัฒนาแบบสอบ ได้หาค่าความตรงตามวิธีการของคาร์เวอว์ หาค่าความเที่ยงโดยใช้สูตร $c = \frac{x^2}{x^2 + N}$

เมื่อ
$$x^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}$$
 โดยที่ b คือ จำนวนผู้สอบ

ผ่าน ครั้งที่ 1 และ ครั้งที่ 2 c คือจำนวนผู้สอบไม่ผ่านทั้งสองครั้ง a คือจำนวนผู้สอบที่สอบผ่านครั้งที่ 1 แต่สอบไม่ผ่านครั้งที่ 2 d คือจำนวนผู้สอบผ่านครั้งที่ 2 แต่สอบไม่ผ่านครั้งที่ 1 และ N คือจำนวนผู้เข้าสอบทั้งหมด

ปี 2526 เช่นกัน มีณี เดือนน้อย (2526) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์เรื่อง "ร้อยละ ชั้นประถมศึกษาปีที่ 4" โดยศึกษากับนักเรียนโรงเรียนวัดคอกไม้ เขตยานนาวา กรุงเทพมหานคร จำนวน 130 คน แบ่งแบบสอบออกเป็น 4-ฉบับย่อย คือ แบบทดสอบความหมายของร้อยละ แบบทดสอบโจทย์ปัญหาร้อยละที่มีจำนวนหนึ่งเป็น 100 หรือ ผลคูณของร้อยละ แบบทดสอบโจทย์ปัญหาร้อยละที่มีคำตอบเป็นจำนวนเต็ม แบบทดสอบโจทย์ปัญหาร้อยละที่มีคำตอบเป็นจำนวนร้อยละ การหาค่าความตรงของแบบสอบใช้วิธีการของคาร์เวอว์ และหาค่าความตรงตามสภาพเมื่อใช้คะแนนเฉลี่ยสะสม เป็นตัวเกณฑ์ หาค่าความ

เที่ยงโดยวิธีการทดสอบซ้ำและหาค่าคตินี้แคป้า สุโขทัย สันตติวงศ์ไชย (2526) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงเกณฑ์วิชาคณิตศาสตร์เรื่อง สมการควอดรติก ชั้นมัธยมศึกษาปีที่ 3" โดยใช้นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนบางบัววิทยาคม จังหวัดสมุทรปราการ จำนวน 160 คน แบ่งแบบสอบออกเป็น 6 ฉบับ คือ การแยกตัวประกอบของโพลิโนเมียล ดีกรีสอง 4 ฉบับ การแก้สมการควอดรติกและการแก้โจทย์สมการควอดรติกอีกอย่างละ 1 ฉบับ โดยหาค่าความตรงของแบบสอบตามวิธีการของคาร์เวอว์ และหาค่าความตรงตามสภาพ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบสอบกับคะแนนเฉลี่ยผลการเรียนวิชาคณิตศาสตร์ของยูเรียน การหาค่าความเที่ยงใช้สูตร
$$C = \sqrt{\frac{X^2}{X^2 + N}}$$

และในปี 2526 เชนกัน เกศริน บุญเกิด (2526) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงเกณฑ์วิชาภาษาไทย เรื่องการเขียนสะกดคำ ชั้นประถมศึกษาปีที่ 4" โดยศึกษากับนักเรียนชั้นประถมศึกษาปีที่ 4 โรงเรียนวัดท่าไทร จังหวัดสุราษฎร์ธานี จำนวน 134 คน แบ่งแบบสอบออกเป็น 3 ฉบับย่อย คือ การเขียนคำสะกดที่เป็นคำใหม่ในบทที่ 19 20 และ 21 ตามลำดับ การหาค่าความตรงใช้วิธีการของคาร์เวอว์ และหาค่าความเที่ยงโดยวิธีการทดสอบซ้ำ และวิธีการของฮวาน

ในปี 2527 ประเทือง ทาสีแสง (2527) ได้ทำวิทยานิพนธ์เรื่อง "การสร้างแบบทดสอบอิงโคเมน เรื่องคู่ลำดับและกราฟ ชั้นมัธยมศึกษาปีที่ 1" โดยศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 1 โรงเรียนประจักษ์ศิลปาคาร จำนวน 167 คน ในปีการศึกษา 2526 ลักษณะของการสร้างข้อสอบเป็นการสร้างขึ้นตามรูปแบบเฉพาะประจำข้อสอบที่กำหนดขึ้น แบ่งแบบสอบออกเป็น 6 ฉบับย่อย คือ แบบสอบฉบับที่ 1 และฉบับที่ 2 เป็นแบบสอบคู่ขนานวัดเนื้อหาคำนวณสามารถในการเข้าใจเรื่องคู่ลำดับ แบบสอบฉบับที่ 3 และฉบับที่ 4 เป็นแบบสอบคู่ขนานวัดในเนื้อหาคำนวณเข้าใจในเรื่องภาพ แบบสอบฉบับที่ 5 และฉบับที่ 6 เป็นแบบสอบคู่ขนานวัดในเนื้อหา เรื่องการเขียนกราฟและหาค่าตอบ การหาค่าความตรงใช้วิธีการให้ยูเชี่ยวชาญด้านเนื้อหาพิจารณา การหาค่าความเที่ยงใช้วิธีการของสวานินาธาน

เมื่อพิจารณาจากวิทยานิพนธ์ที่ศึกษา เกี่ยวกับแบบสอบอิงเกณฑ์ซึ่งได้กล่าวมาแล้วทั้งหมด มีประเด็นที่เป็นข้อสังเกต 6 ประการ คือ

1. งานค้นคว้าแทบทั้งหมดจะยังคง เป็นเพียงการสร้างแบบสอบอิงเกณฑ์ขึ้นมาตามหลักการของการสร้างแบบสอบอิงเกณฑ์เท่านั้น โดยเนื้อหาส่วนใหญ่ที่นำมาสร้างจะเป็นเนื้อหาในวิชา

คณิตศาสตร์ งานของกาญจน วัชรสุนทร นอกจากจะจัดเป็นงานวิจัยชิ้นแรกในขณะนี้แล้ว ยังเป็นงานวิจัยชิ้นแรกที่ใช้รูปแบบเฉพาะประจำข้อสอบในการสร้างข้อสอบขึ้นใช้ในแบบสอบ ในขณะที่คนอื่นยังคงยึดเอาวัตถุประสงค์เชิงพฤติกรรมเป็นหลักในการสร้างข้อสอบ เพิ่งจะมีปรากฏอีกครั้งในปี 2527 ที่ประเทือง ทาสีแสง ได้พัฒนาแบบสอบอิงโคเมนขึ้นมาอีกครั้งหนึ่ง สำหรับการค้นคว้าในทำนองอื่น ๆ นอกเหนือจากการสร้างแบบสอบ ก็ยังมีงานของประภา แก่นเพิ่ม และ ชมพู่ จันทร์อมรพร ที่มุ่งหามาตรการในการกำหนดจุดตัด

2. การวิเคราะห์ค่าประจำข้อสอบ งานทั้งหมดที่ศึกษา ยังคงยึดเอาค่าความยาก และการหาค่าดัชนีเอส ซึ่งยังคงเรียกว่า "อำนาจจำแนก" แม้ว่าในระยะหลังจะปรากฏชื่อว่า รอยคและฮาลาโกนาจะใช้ชื่อให้เฉพาะลงไปสำหรับการวิเคราะห์กับแบบสอบอิงเกณฑ์ว่า "ดัชนีความไวในการสอน" (Instructional Sensivity Index)

3. การกำหนดเกณฑ์มาตรฐาน โดยทั่วไป การกำหนดเกณฑ์มาตรฐานให้กับแบบสอบที่สร้างขึ้นยังคงยึดเอาแนวคิดของเบอร์กเป็นวิธีการ จะมีวิธีอื่นบ้างก็คือการใช้เกณฑ์ภายนอกเข้ามาเป็นเครื่องพิจารณา กำหนดเกณฑ์มาตรฐาน เกณฑ์ภายนอกที่ใช้กันมากคือ คะแนนผลการเรียนในรายวิชาต่างๆ ที่เกี่ยวข้องกับแบบสอบที่พัฒนาขึ้น

4. การหาค่าความตรงของแบบสอบ นอกจากการหาค่าความตรงตามเนื้อหาโดยทั่วไปแล้ว ส่วนใหญ่ก็ยังคงใช้วิธีการหาค่าความตรงโดยวิธีการของคาร์เวอว์ ซึ่งแนวคิดนี้จะสอดคล้องกับวิธีการกำหนดเกณฑ์มาตรฐานของเบอร์ก คือ ถ้าใช้วิธีการกำหนดเกณฑ์มาตรฐานของเบอร์กแล้ว ณ จุดเกณฑ์มาตรฐานที่กำหนดขึ้นจะมีค่าความตรงตามวิธีการของคาร์เวอว์สูงที่สุดควบคู่กัน

5. การหาค่าความเที่ยงของแบบสอบ ส่วนใหญ่จะยังใช้การหาค่าความคงที่ในการจัดประเภทระดับการรับรู้ของผู้สอบ โดยอาจจะใช้สูตรในการคำนวณต่างกันไป เช่น วิธีการของสวานินาธาน, สับโคเวียค หรือ ฮวน มีงานวิจัยของกาญจน วัชรสุนทร และ ประภา แก่นเพิ่ม ที่ใช้วิธีการหาค่าความเที่ยงโดยใช้สูตรของลิฟวิงสตัน

6. การกำหนดความยาวของแบบสอบ ในประเด็นนี้กล่าวได้ว่า ในการพัฒนาแบบสอบอิงเกณฑ์เท่าที่นำมา ยังไม่ได้มีใครได้คำนึงถึงปัญหาข้อแต่อย่างใด การกำหนดความยาวที่ทำกันจึงยังคงเป็นเพียงการกำหนดขึ้นมาตามที่เห็นสมควร มากกว่าที่จะยึดหลักวิชาใดๆ

จากงานเขียนที่เกี่ยวข้องซึ่งได้กล่าวนำแล้วทั้งหมด จะเห็นว่าในช่วงเวลาประมาณ 20 ปีที่ผ่านมา แนวคิดเชิงหลักวิชาต่างๆ เกี่ยวกับแบบสอบอิงเกณฑ์ได้รับการพัฒนาขึ้นมามาก แต่ในขณะที่เดียวกันก็ยังคงมีปัญหา ซึ่งวิชาการที่กองการการศึกษาค้นคว้าอีกหลายประเด็น เช่นกัน ยกเว้นในประเด็นเรื่อง ความยาวของแบบสอบอิงเกณฑ์ที่มุ่งศึกษาในครั้งนี้ ก็ยังคงมีปัญหาคืออีกหลายประเด็นที่ยังไม่มีการศึกษาค้นคว้า เช่น

1. วิธีการที่จะกำหนดขอบเขตของเนื้อหา (Domain Specification) ให้ชัดเจน วิธีการสร้างและสุ่มข้อสอบให้มีความเป็นตัวแทนของขอบเขตเนื้อหาที่กำหนด ทั้งนี้ อาจเป็นไปได้ว่า ลักษณะของเนื้อหาวิชาที่ต่างกัน อาจจะต้องใช้วิธีการกำหนดขอบเขตของเนื้อหาต่างกัน วิธีการที่ใช้ในการสร้างและสุ่มข้อสอบก็อาจจะต้องใช้วิธีการที่ต่างกันออกไปด้วย ในปัจจุบันการสร้างแบบสอบอิงเกณฑ์ส่วนใหญ่จะยังคงทำกับวิชาคณิตศาสตร์เท่านั้น แต่ไม่ค่อยปรากฏการพัฒนาแบบสอบอิงเกณฑ์ในเนื้อหาวิชาอื่นมากนัก โดยเฉพาะวิชาที่มีโครงสร้างของเนื้อหาแตกต่างไปจากวิชาคณิตศาสตร์

2. แนวคิดการกำหนดเกณฑ์มาตรฐาน แม้ในระยะหลังๆค่อนข้างจะปรากฏออกมาว่าวิธีการกำหนดเกณฑ์มาตรฐานต่างกัน จะทำให้ได้คะแนนเกณฑ์มาตรฐานที่ต่างกัน และคงไม่มีวิธีการใดที่ดีที่สุด แต่ในขณะที่เดียวกันก็ยังคงขาดการศึกษาค้นคว้าอีก เช่นกันว่า ในสถานการณ์การสอบเนื้อหาที่สอบ และกลุ่มผู้สอบซึ่งต่างกันออกไปในลักษณะใดน่าจะมีมาตรการใดในการดำเนินการกำหนดเกณฑ์มาตรฐาน เหมือนหรือแตกต่างกันอย่างไร

3. ปัญหาการกำหนดมาตรการวิธีการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์ที่เหมาะสม ทั้งนี้ อาจจะเป็นไปได้ว่า ในลักษณะของการทดสอบที่น่าจะแน่นอน ไปใช้เพื่อการตัดสินใจที่ต่างกันนั้น มาตรการในการหาค่าความเที่ยง ตลอดจนนิยามของคำว่า "ความเที่ยง" อาจจะต้องมีความแตกต่างกันออกไปด้วย

4. มาตรการของการหาค่าความตรงประเภทอื่นนอกเหนือไปจากความตรงตามเนื้อหาแบบสอบอิงเกณฑ์ที่พัฒนาขึ้นมา นั้น ถ้าหากผู้พัฒนาต้องการที่จะนำไปใช้เพื่อการใดโดยเฉพาะ เช่น การระบุจุดบกพร่องของผู้เรียน ใช้เพื่อคัดเลือกบุคคลที่รอบรู้ไปทำงานหรือกิจกรรมใดๆ หรือสร้างขึ้นมาโดยมีหลักการหรือทฤษฎีใดๆ โดยเฉพาะแล้ว ก็ควรจะต้องมีความจำเป็นที่จะต้องดำเนินการหาค่าความตรงที่เหมาะสมกับแบบสอบอิงเกณฑ์เฉพาะที่สร้างขึ้นมากด้วย ปัจจุบันการศึกษาในประเด็นนี้ นับได้ว่ามีอยู่น้อยมาก

5. ในประเด็นสุดท้าย การศึกษาค้นคว้าต่างๆ เกี่ยวกับแบบสอบถามเชิงคุณภาพในปัจจุบัน ยังคงดำเนินการโดยยึดเอาแบบสอบถามที่พัฒนาขึ้นมาได้ เป็นแกนกลางของการค้นคว้า แต่อาจจะ เป็นไปไกว่า การศึกษาค้นคว้าในระยะต่อไปอาจจะมี ความจำเป็นที่จะต้องศึกษาถึงผลที่ เกิดขึ้น จากการทำเอาแบบสอบถามเชิงคุณภาพที่พัฒนาขึ้นไป ไขกับกลุ่มผู้เรียนที่มีภูมิหลังต่างๆแตกต่างกัน

ประเด็นต่างๆ เหล่านี้ เป็นสิ่งที่ต้องการ ผู้ที่จะเข้ามาศึกษาค้นคว้าอีกมาก ผู้ที่สนใจศึกษา ควรจะ เริ่มต้นด้วยการอ่านบทความต่างๆ เกี่ยวกับแบบสอบถามเชิงคุณภาพที่มีอยู่ให้ เข้าใจอย่างถ่องแท้ เพื่อที่จะสามารถเข้าใจประเด็นปัญหาในแต่ละข้อได้อย่างลึกซึ้ง แล้วจึงค่อยวางแผนการดำเนินการ การวิจัยในรายละเอียด

2. เทคนิคการวิเคราะห์ความล่าช้า

ดังที่กล่าวมาแล้วในบทนำว่า เทคนิคการวิเคราะห์ความล่าช้าขั้นพัฒนาขึ้นมาโดยนัก-คณิตศาสตร์ชื่อ แอบราฮัม วอลค์ ในปี 1945 เทคนิคนี้มุ่งจะตรวจสอบเพื่อตัดสินว่าจะยอมรับ สมมุติฐานศูนย์ (Null Hypothesis) หรือสมมุติฐานอื่น (Alternative Hypothesis) ภายใต้โอกาสความคลาดเคลื่อนที่กำหนดขึ้นไว้ล่วงหน้า โดยในการทดสอบจะมีการสุ่มตัวอย่างขึ้น มาทีละหน่วย แล้วพิจารณาว่าจากข้อมูลที่ได้ ควรจะตัดสินใจรับสมมุติฐานใด ถ้ายังไม่สามารถจะ ตัดสินใจได้ ก็จะสุ่มตัวอย่างเพิ่ม เข้าไปอีกทีละหน่วย เมื่อตัดสินใจยอมรับสมมุติฐานตัวใดตัวหนึ่ง ได้ก็จะยุติการทดสอบ จุดเด่นของหลักการนี้คือต้องการใช้กลุ่มตัวอย่างให้น้อยที่สุดในการตัดสินใจ ยอมรับสมมุติฐานตัวใดตัวหนึ่ง โดยมีโอกาสเกิดความคลาดเคลื่อนในการตัดสินใจในระดับต่ำ ซึ่ง จะกำหนดขึ้นไว้ล่วงหน้าก่อนที่จะดำเนินการทดสอบ

แนวคิดดังกล่าวนี้ ได้ถูกพัฒนาขึ้นใช้ในโรงงานอุตสาหกรรม ดังตัวอย่างแนวคิดทาง คณิตศาสตร์ที่จะอธิบายต่อไปนี้ เป็นการมุ่งใช้เทคนิคการวิเคราะห์ความล่าช้า เพื่อตรวจสอบค่า มีชนิดมี เลขคณิตของตัวแปรที่มีลักษณะการแจกแจง เป็นแบบทวินาม

บริษัทแห่งหนึ่งผลิตสินค้าออกจำหน่าย ลักษณะของสินค้าบรรจุอยู่ในกล่อง บริษัทถือว่า สินค้าที่ผลิตได้มีความสมบูรณ์และส่งออกจำหน่ายได้ ถ้าสัดส่วนของสินค้าในกล่องเสียไม่ เกินอัตรา P_0 และจะถือว่าสินค้าที่ผลิตได้มีความบกพร่อง ไม่สามารถส่งออกจำหน่ายได้ ถ้าสัดส่วนของ สินค้าในกล่องเสีย เกินอัตรา P_1 ในการตรวจสอบคุณภาพสินค้าดังกล่าว ไม่สามารถตรวจดูสินค้า ทุกๆชิ้นได้ ปัญหาจึงมีอยู่ว่า จะสุ่มสินค้าขึ้นมาตรวจสอบโดยให้มีจำนวนน้อยชิ้นที่สุดและได้ข้อสรุป ที่มีความผิดพลาดไม่ เกินอัตราที่กำหนดได้อย่างไร

ในการทดสอบ สมมติให้สัดส่วนของสินค้าที่เสียจริงซึ่งบรรจุอยู่ในกล่อง เป็น
คั้งแล้วในการทดสอบ

จะยอมรับ H_0 ก็ต่อเมื่อ $P \leq P_0$

และยอมรับ H_a ก็ต่อเมื่อ $P \geq P_1$

ทั้งนี้กำหนดให้ α เป็นโอกาสของความคลาดเคลื่อนในการปฏิเสธ H_0
ที่เป็นจริง

β เป็นโอกาสของความคลาดเคลื่อนในการปฏิเสธ H_1
ที่เป็นจริง

$1 - \alpha$ เป็นโอกาสของความถูกต้องในการยอมรับ H_0 ที่เป็นจริง

$1 - \beta$ เป็นโอกาสของความถูกต้องในการยอมรับ H_a ที่เป็นจริง

ในการดำเนินการทดสอบ ถ้ามีการสุ่มสินค้าขึ้นมา m ชิ้น และ d_m เป็นจำนวนสินค้า
ที่เสีย จะได้ว่า

กรณีที่ H_0 เป็นจริง โอกาสที่จะเกิดขึ้นก็จะเป็น $P_0^{d_m} (1 - P_0)^{m - d_m}$

กรณีที่ H_a เป็นจริง โอกาสที่จะเกิดขึ้นก็จะเป็น $P_1^{d_m} (1 - P_1)^{m - d_m}$

ในการวิเคราะห์ตามลำดับชั้น จะมีการตรวจสอบอัตราส่วนความน่าจะเป็น

$$\frac{P_1^{d_m} (1 - P_1)^{m - d_m}}{P_0^{d_m} (1 - P_0)^{m - d_m}}$$

ว่าอยู่ในลักษณะเช่นไร โดยถ้า

$$\frac{P_1^{d_m} (1 - P_1)^{m - d_m}}{P_0^{d_m} (1 - P_0)^{m - d_m}} < \frac{1 - \beta}{\alpha}$$

ดังนั้น จะมีความหมายว่า โอกาสที่ H_a จะเป็นจริงต่อโอกาสที่ H_0 จะเป็นจริง
จะมีค่าน้อยกว่าอัตราส่วนความเชื่อมั่นที่จะยอมรับ H_a ที่ถูกต่ออัตราส่วนที่จะยอมรับ H_0 ที่ผิด
กรณีเช่นนี้ก็จะยอมรับว่า H_a เป็นจริง

ถ้าผลการทดสอบปรากฏออกมาว่า

$$\frac{P_1^{d_m}(1 - P_1)^{m - d_m}}{P_0^{d_m}(1 - P_0)^{m - d_m}} > \frac{\beta}{1 - \alpha}$$

จะมีความหมายว่า โอกาสที่ H_a เป็นจริงต่อโอกาสที่ H_0 จะเป็นจริงมากกว่า
โอกาสที่จะยอมรับ H_a ที่ผิด และยอมรับ H_0 ที่ถูก กรณีเช่นนี้ก็จะยอมรับ H_0

ถ้าผลการทดสอบปรากฏอยู่ในลักษณะ

$$\frac{\beta}{1 - \alpha} < \frac{P_1^{d_m}(1 - P_1)^{m - d_m}}{P_0^{d_m}(1 - P_0)^{m - d_m}} < \frac{1 - \beta}{\alpha}$$

ก็จะต้องมีการสุ่มตัวอย่างเพิ่ม และดำเนินการทดสอบต่อ

จากสมการข้างบน ที่ถ้าสามารถทราบค่า หรือปรับสมการให้อยู่ใต้ตามลำพังแล้ว
จะสามารถแปลความหมายได้ง่ายขึ้น คือตัว d_m จึงได้มีการใช้ Logarithm เข้าช่วย
โดยจะเป็นไปตามกระบวนการดังนี้

$$\log \frac{\beta}{1 - \alpha} < \log \frac{P_1^{d_m}(1 - P_1)^{m - d_m}}{P_0^{d_m}(1 - P_0)^{m - d_m}} < \log \frac{1 - \beta}{\alpha}$$

$$\log \frac{\beta}{1 - \alpha} < d_m \log P_1 + (m - d_m) \log(1 - P_1) - d_m \log P_0$$

$$+ (m - d_m) \log(1 - P_0) < \log \frac{1 - \beta}{\alpha}$$

$$\log \frac{\beta}{1-\alpha} d_m \left(\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0} \right) - m \log \frac{1-P_1}{1-P_0} < \log \frac{1-\beta}{\alpha}$$

$$\frac{\log \frac{\beta}{1-\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + \frac{m \left(\log \frac{1-P_1}{1-P_0} \right)}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} < d_m < \frac{m \left(\log \frac{1-P_1}{1-P_0} \right)}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}} + \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}}$$

เพื่อให้สมการอยู่ในรูปที่ง่ายเข้า
กำหนดให้

$$S = \frac{\log \frac{1-P_1}{1-P_0}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}}$$

$$h_o = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}}$$

และ

$$h_a = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{P_1}{P_0} - \log \frac{1-P_1}{1-P_0}}$$

จะได้ว่า

$$h_o + ms < d_m < h_a + ms$$

$$L_1 = h_1 + ms$$

$$L_o = h_o + ms$$

เมื่อกำหนดให้ L_1 เป็นเส้นสมการที่แยกบริเวณสินค้าไม่ได้มาตรฐานออกจากเขตทดสอบคอ

L_2 เป็นเส้นสมการที่แยกบริเวณสินค้าได้มาตรฐานออกจากเขตทดสอบคอ

โดยที่ถ้านำมาเขียนเป็นเส้นกราฟแล้ว h_1 และ h_o จะเป็นจุดตัดกับแกน และ s จะเป็นค่าความชันของเส้นตรงสองเส้นซึ่งขนานกัน และการที่จะสร้างเส้นกราฟไถ่นั้น จะต้องกำหนดค่าต่างๆขึ้น 4 ค่า คือ ค่าพารามิเตอร์เมื่อ h_o เป็นจริง ค่าพารามิเตอร์เมื่อ h_a เป็นจริง อัตราความคลาดเคลื่อนประเภทที่หนึ่งและอัตราความคลาดเคลื่อนประเภทที่สอง

ดังนี้ ถ้าลองกำหนดให้

$$\alpha = .05 \quad \beta = .05$$

$$P_o = .20 \quad P_1 = .45$$

จะได้ค่า

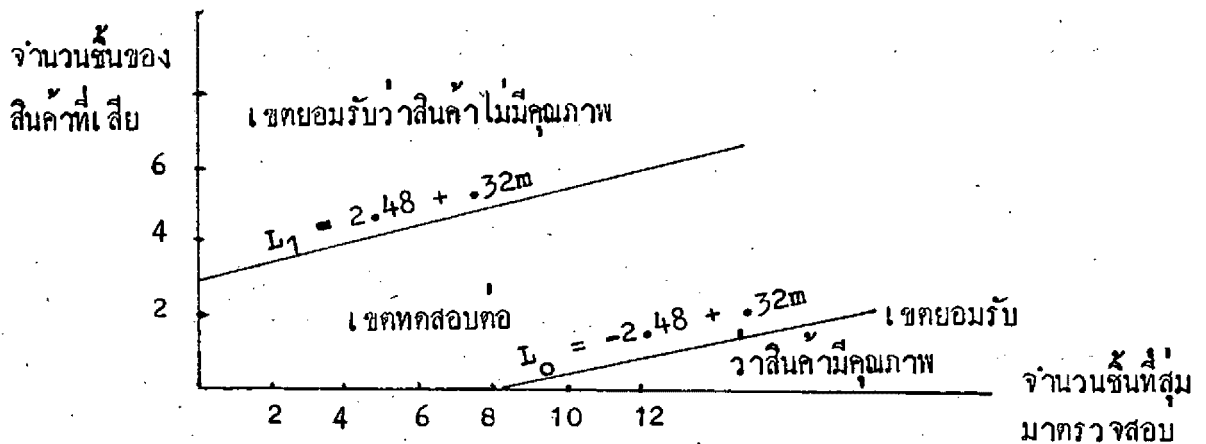
$$h_1 = 2.48$$

$$h_o = -2.48$$

$$s = 0.32$$

ซึ่งเมื่อนำมาเขียนเป็นเส้นกราฟจะได้ดังนี้

แผนภูมิที่ 7 การสร้างเส้นกราฟของเทคนิคการวิเคราะห์ตามลำดับชั้นเพื่อพิจารณาทัศนียภาพยอมรับสมมุติฐานตัวใดตัวหนึ่ง



ดังนั้น จะถือว่า ถ้าสุ่มสินค้าขึ้นมา 8 ชั้น
 จะถือว่าสินค้าไม่มีคุณภาพ ถ้ามีสินค้าเสียถึง 5 ใน 8 ชั้น
 จะถือว่าสินค้ามีคุณภาพ ถ้าไม่มีสินค้าใดเสียเลย
 ถ้าสินค้าเสียอยู่ในช่วง 1 - 4 ชั้น จะต้องดำเนินการทดสอบต่อ
 ถ้าสุ่มสินค้าขึ้นมา 12 ชั้น
 จะถือว่าสินค้าไม่มีคุณภาพ ถ้าเสียถึง 6 ชั้นขึ้นไปใน 12 ชั้น
 จะถือว่าสินค้ามีคุณภาพ ถ้าเสียเพียง 1 ใน 12 ชั้น
 ถ้าสินค้าเสีย 2 - 5 ชั้น จะถือว่าต้องทดสอบต่อ

การประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นในทางการศึกษา

จากการพิจารณาความต่างๆในทางการศึกษาที่มุ่งประยุกต์เทคนิควิเคราะห์ตามลำดับชั้นมาใช้เท่าที่คิดว่า สามารถที่จะแบ่งลักษณะของการประยุกต์ได้ตามช่วงเวลาและพัฒนาการของการประยุกต์ได้เป็น 3 ระยะ คือ

1. การประยุกต์ในระยะแรกเริ่ม
2. การประยุกต์ในระยะของการพัฒนาเพิ่มเติมในรายละเอียดของหลักวิชา
3. การประยุกต์ใช้กับการทดสอบในระบอบอิงเกณฑ์

1. การประยุกต์ในระยะแรก เริ่ม

การประยุกต์ในระยะนี้ปรากฏในปี 1946 - 1960 โดยมีลักษณะเด่นๆ คือ

ก. ในบทความที่นำเสนอผลการศึกษาโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น จะมีการแนะนำแนวคิดในหลักวิชาทางด้านสถิติของเทคนิคการวิเคราะห์ตามลำดับชั้นมาก่อน

ข. ลักษณะของการประยุกต์ เป็นการใช้เทคนิคการวิเคราะห์ตามลำดับชั้นอย่างตรงไปตรงมา ยังไม่มีการดัดแปลงหรือพัฒนาเพิ่มเติมในหลักวิชาทางสถิติ

บทความเกี่ยวกับการประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นในทางการศึกษาระดับแรกเท่าที่ค้นคว้าได้คืองานของโคว์เดน (Cowden) ในปี 1946 โดยโคว์เดนได้นำเทคนิคการวิเคราะห์ตามลำดับชั้นมาศึกษาผลการจัดประเภทบุคคลตามระดับการรอบรู้ โดยใช้ นักศึกษาจำนวน 10 คน ที่ศึกษาวิชาสถิติพื้นฐานในมหาวิทยาลัยนอร์ธคาโรไลนา แบบสอบเต็มฉบับที่ใช้มีความยาว 200 ข้อ การประยุกต์มุ่งตรวจสอบผลการตัดสินใจจัดประเภทบุคคลตามระดับการรอบรู้ที่ช่วงความยาว 20 ข้อ 40 ข้อ 60 ข้อ และ 80 ข้อ ตามลำดับ โดยกำหนดเกณฑ์มาตรฐานในการให้ผ่านเป็น 0.30 และเกณฑ์มาตรฐานในการให้ตกเป็น 0.40 อัตราความคลาดเคลื่อนในการให้ผู้นับรอบรู้สอบตกเป็น 0.20 และอัตราความคลาดเคลื่อนในการให้ผู้นับรอบรู้สอบผ่านเป็น 0.10

ประเด็นหลักของข้อค้นพบ คือ เทคนิคการวิเคราะห์ตามลำดับชั้นสามารถที่จะใช้ตัดสินผลเพื่อจัดประเภทบุคคลตามระดับการรอบรู้ของผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนสูงหรือต่ำมากๆ ได้อย่างรวดเร็ว ในขณะที่ผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนในระดับปานกลางจะต้องใช้จำนวนข้อสอบมากกว่า เพื่อที่จะรับระดับความสามารถ

ในปี 1950 มูนาน (Moonan) ได้ศึกษาลักษณะของคะแนนที่ได้จากการใช้เทคนิคการวิเคราะห์ตามลำดับชั้น กับคะแนนจากแบบสอบเต็มฉบับ โดยใช้ นักศึกษาจำนวน 39 คนที่เรียนวิชาสถิติภาคบรรยายในมหาวิทยาลัยมิเนโซต้า แบบสอบที่ใช้ประกอบด้วยข้อสอบจำนวน 75 ข้อ โดยข้อสอบทั้งหมดในแบบสอบได้รับการเรียงโดยใช้การสุ่มอย่างง่าย ในการศึกษา มูนานได้กำหนดค่าเกณฑ์มาตรฐานในการให้ผ่าน (P_0) เกณฑ์มาตรฐานในการให้ตก (P_1) อัตราความคลาดเคลื่อนในการให้ผู้นับรอบรู้สอบตก (α) อัตราความคลาดเคลื่อนในการให้ผู้นับรอบรู้สอบผ่าน (β) แตกต่างกัน 5 ลักษณะ คือ

1.	$P_0 = 35/75$	$P_1 = 45/75$	$\alpha = .10$	$\beta = .10$
2.	$P_0 = 25/75$	$P_1 = 45/75$	$\alpha = .02$	$\beta = .02$
3.	$P_0 = 25/75$	$P_1 = 40/75$	$\alpha = .05$	$\beta = .05$
4.	$P_0 = 30/75$	$P_1 = 40/75$	$\alpha = .10$	$\beta = .10$
5.	$P_0 = 35/75$	$P_1 = 45/75$	$\alpha = .10$	$\beta = .20$

ผลการศึกษาพบว่า คะแนนที่ได้จากการทดสอบตามเทคนิคการวิเคราะห์ตามลำดับชั้น และคะแนนจากแบบสอบเติมฉบับมีค่าสัมประสิทธิ์สหสัมพันธ์โดยเฉลี่ยสูง 0.90 แต่จากการวิเคราะห์ความแปรปรวนพบว่า ค่าคะแนนเฉลี่ยที่ได้หลังจากปรับฐานคะแนนให้เท่ากับคะแนนจากแบบสอบเติมฉบับแล้วแตกต่างกันไปจากค่าคะแนนเฉลี่ยจากแบบสอบเติมฉบับอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

ในปี 1958 ฟิสก์และโจน (Fiske and Jones) ได้เขียนบทความเผยแพร่และสนับสนุนการใช้เทคนิคการวิเคราะห์ตามลำดับชั้นกับการศึกษาทางจิตวิทยา โดยได้เสนอแนวคิดและหลักวิชาทางสถิติ พร้อมทั้งให้ข้อเสนอแนะว่า ในกรณีของการศึกษาที่การเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจะต้องสิ้นเปลืองค่าใช้จ่ายมากแล้ว เทคนิคการวิเคราะห์ตามลำดับชั้นจะมีบทบาทเข้ามาช่วยในการลดค่าใช้จ่ายลงได้ เพราะกลุ่มตัวอย่างที่จำเป็นจะต้องใช้ในการศึกษาสามารถลดลงไปไ้มาก

2. การประยุกต์ในระยะของการพัฒนาเพิ่มเติมในรายละเอียดของหลักวิชา

การประยุกต์ในช่วงนี้ เริ่มมีปรากฏบทความในปี 1968 - 1972 บทความในช่วงนี้มีลักษณะที่สำคัญ คือ

ก. ในการนำเสนอบทความจะไม่มีกรนำเสนอแนวคิดหลักวิชาต่างๆ ในประเด็นของหลักการทางสถิติของเทคนิคการวิเคราะห์ตามลำดับชั้น ดังเช่นที่เคยปรากฏในระยะแรก

ข. การศึกษาจะมีลักษณะของการมุ่งเน้นเฉพาะลงไป ซึ่งปรากฏชัดในลักษณะย่อย

3. ประการ คือ

1. ศึกษาในเชิงทฤษฎี เช่น ผลการจับประเภทบุคคลตามระดับการรอบรู้จากประชากรที่มีการกระจายในรูปแบบต่างๆ

2. ผลการประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้น เปรียบเทียบกับการใช้เทคนิคอื่นๆ
3. การพัฒนาหลักการทางสถิติของเทคนิคการวิเคราะห์ตามลำดับชั้นให้มีรายละเอียดมากขึ้น

งานวิจัยในช่วงนี้ก็มีผลการวิจัยของแพทเทอร์สัน (Patterson) ที่อ้างถึงในงานวิจัยของเคลียร์รี, ลินน์ และ รอค (Cleary, Linn and Rock) ในปี 1968 โดยแพทเทอร์สันได้ศึกษาผลการจัดประเภทบุคคลตามระดับการรอบรู้ เมื่อใช้ข้อมูลจากการกำหนดชั้นเพื่อให้ได้ประชากรที่ตรงการศึกษามีการกระจายในลักษณะต่างๆ ซึ่งจากผลการศึกษาพบว่าไม่ว่าลักษณะการแจกแจงของข้อมูลในประชากรจะเป็นอย่างไร เทคนิคการวิเคราะห์ตามลำดับชั้นสามารถที่จะจัดประเภทบุคคลที่มีความสามารถสูงสุดและต่ำสุดได้ถูกต้องโดยใช้ข้อสอบเพียงน้อยข้อ

เคลียร์รี, ลินน์ และ รอค (Cleary, Linn and Rock 1968) ได้ศึกษาผลการจัดประเภทบุคคลตามระดับการรอบรู้ โดยแยกความสามารถออกเป็น 4 ระดับ เมื่อใช้วิธีการต่างๆ 4 วิธี คือ การทดสอบที่เป็นลักษณะสองชั้นก่อน การจัดสอบจากข้อสอบที่มีค่าความยากแตกต่างกันมาก การใช้ผลความแตกต่างจากการสอบภายในกลุ่ม และการใช้เทคนิคการวิเคราะห์ตามลำดับชั้น ซึ่งจากการศึกษาพบว่า เทคนิคการวิเคราะห์ตามลำดับชั้นสามารถจัดประเภทบุคคลตามระดับการรอบรู้ได้ถูกต้องสูงสุด คือ จัดประเภทบุคคลได้ถึง 1,750 คน ในจำนวนกลุ่มตัวอย่างทั้งหมดที่แยกไว้เพื่อการตรวจสอบ (Cross Validation Sample) จำนวน 2,407 คน คิดเป็นร้อยละ 72.70

ในปี 1972 ลินน์, รอค และ เคลียร์รี (Linn, Rock and Cleary) ได้เขียนบทความชี้แนะการใช้เทคนิคการวิเคราะห์ตามลำดับชั้นที่ได้รับการพัฒนาเพิ่มเติมในหลักการทางสถิติคือ

- ก. การกำหนดค่า P_0 และ P_1 เปลี่ยนจากแนวคิดของการคำนวณจากจำนวนข้อที่ทำได้ มาใช้จำนวนข้อที่ทำได้แทน
- ข. นำเอาค่าความยากของข้อสอบที่ได้จากการสอบมารวมในการคำนวณ
- ค. สามารถใช้ได้กับภาวะการตัดสินใจจัดประเภทบุคคลออกเป็น 2 ส่วน คือ ผู้รอบรู้และผู้ไม่รอบรู้

ทั้งนี้ มีหลักการทางสถิติ คือ

ในการทดสอบ จะจัดผู้สอบเป็นผู้รอบรู้และยุติการทดสอบเมื่อ

$$(m) = \sum_{i=1}^m \log R_i > A$$

โดยที่ค่าผู้สอบทำข้อสอบข้อนั้นๆ ถูก

$$R_i = \frac{P_{h_i}}{P_{l_i}}$$

และค่าผู้สอบตอบข้อสอบนั้นๆ ผิด

$$R_i = \frac{(1 - P_{h_i})}{(1 - P_{l_i})}$$

P_{h_i} คือค่าความยากของข้อสอบในกลุ่มสูง

P_{l_i} คือค่าความยากของข้อสอบในกลุ่มต่ำ

$$A = \frac{1 - \alpha}{\beta}$$

m คือจำนวนข้อสอบทั้งหมดที่ได้รับ การทดสอบ

α คืออัตราความคลาดเคลื่อนในการให้ผู้รอบรู้สอบผิด

β คืออัตราความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน

ในการทดสอบจะจัดผู้สอบเป็นผู้ไม่รอบรู้ และยุติการทดสอบเมื่อ

$$(m) = \sum_{i=1}^m \log R_i < -A$$

โดยที่

$$A = \frac{\alpha}{1 - \beta}$$

ทั้งนี้ในการทดสอบในการศึกษาเฉพาะกรณีของ ลินน์, รอค และ เคลียร์ ถ้าผู้สอบทำข้อสอบหมดทั้ง 60 ข้อ และยังไม่สามารถจับประเภทระดับการรอบรู้ได้ จะใช้เกณฑ์พิจารณาต่อไป คือ ถ้า

$$\Psi(60) \geq 0$$

จะถือว่าผู้สอบเป็นผู้รอบรู้ และถ้า

$$\Psi(60) < 0$$

จะถือว่าผู้สอบเป็นผู้ไม่รอบรู้

วิธีการที่ ลินน์, รอค และ เคลียร์ พัฒนาขึ้นมาใช้ได้แม้ในกรณีที่ข้อสอบจะมีค่าความยากไม่เท่ากัน ทั้งนี้เพราะผลการสอบข้อสอบแต่ละข้อ ไม่ว่าจะ ถูก หรือ ผิด จะส่งผลกระทบต่อโอกาสในการที่จะจับประเภทบุคคลตามระดับการรอบรู้ กล่าวคือ ถ้าตอบข้อสอบที่ยากได้ถูกต้อง ก็จะมีความเป็นไปได้สูงที่จะได้รับการจับประเภทเป็นผู้รอบรู้ และถ้าตอบข้อสอบที่ง่ายผิด ก็จะมีความเป็นไปได้ที่จะได้รับการจับประเภทเป็นผู้ไม่รอบรู้ แต่อย่างไรก็ตามวิธีการนี้มีความยุ่งยากในการปฏิบัติ เพราะจะต้องมีการพิจารณาตัดสินผลการจับประเภทบุคคลเป็นรายคน และต้องจัดกระทำทุกครั้งที่มีการตอบข้อสอบข้อหนึ่งๆ

จากการใช้เทคนิคที่พัฒนาขึ้นดังกล่าว จับประเภทบุคคลตามระดับการรอบรู้จากผลการสอบในวิชา คณิตศาสตร์ วิทยาศาสตร์ ธรรมชาติ และเรียงความภาษาอังกฤษ เปรียบเทียบกับผลการทดสอบแบบปกติทั่วไป เมื่อกำหนดให้ผลการจับประเภทมีความถูกต้องเท่ากัน ปรากฏว่าการสอบในภาวะปกติจะต้องใช้จำนวนข้อสอบมากกว่าการสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับขั้นถึงประมาณ 2 เท่า ดังรายละเอียดในตารางที่ 9

ตารางที่ 9 เปรียบเทียบสัดส่วนจำนวนข้อสอบที่ต้องใช้ในการทดสอบภาวะปกติกับการทดสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น เพื่อให้ผลความถูกต้องในการจัดประเภทบุคคลตามระดับการรอบรู้มีจำนวนเท่ากัน จากการศึกษาของ ลินน์, รอค และ เคลียร์

อัตราความคลาดเคลื่อน	แบบสอบ		
	คณิตศาสตร์	วิทยาศาสตร์ ธรรมชาติ	เรียงความ ภาษาอังกฤษ
$\alpha = \beta = .25$	2.17	2.67	2.08
$\alpha = \beta = .10$	2.35	1.97	1.95
$\alpha = \beta = .05$	2.69	1.79	1.84
$\alpha = \beta = .01$	2.33	1.65	1.71

(Linn, Rock and Cleary 1972 : 93)

3. การประยุกต์ใช้กับการทดสอบในระบบอิงเกณฑ์

ในช่วงปี 1970 เป็นต้นมา แนวคิดของการทดสอบในระบบอิงเกณฑ์แนวใหม่ได้เริ่มมีการพัฒนาและการกล่าวถึงกันอย่างจริงจัง ใช้นักวิชาการเริ่มกล่าวพาดพิงถึงการประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นมาใช้ในการทดสอบอิงเกณฑ์ เพื่อจัดประเภทบุคคลตามระดับการรอบรู้ โดยใช้ข้อสอบจำนวนน้อยข้อ ดังมีปรากฏในบทความของ เกลเซอร์และนิตโก (Glaser and Nitko in Popham 1971) และมิลล์แมน (Millman 1973) แต่อย่างไรก็ดี บุคคลทั้งสองก็ไม่ได้เสนอแนวคิดในรายละเอียดแก่ประการใด จนกระทั่งปี 1975 สเตคไลน์และพิริยานุวัชน์ (Stecklein and Pitiyanuwat 1975) ได้เขียนบทความสนับสนุนและเสนอวิธีการ ตลอดจนขั้นตอนในการดำเนินงานในการใช้เทคนิคการวิเคราะห์ตามลำดับชั้น เพื่อจัดประเภทบุคคลตามระดับการรอบรู้ ซึ่งในบทความผู้เขียนได้ชี้แนะประเด็นความเหมาะสมในการนำเทคนิคการวิเคราะห์ตามลำดับชั้นมาใช้ในการทดสอบในระบบอิงเกณฑ์ สรุปไว้ดังนี้

ก. การทดสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้นในระบบการทดสอบแบบอิงเกณฑ์ แม้จะใช้ข้อสอบเพียงจำนวนน้อยข้อ ข้อสอบเหล่านี้ก็มีความเป็นตัวแทนของเนื้อหาวิชาที่ถือว่าการทดสอบ ทั้งนี้ เนื่องจากว่าแบบสอบอิงเกณฑ์จะมีการกำหนดขอบเขตของเนื้อหาให้มีความเป็นเอกพันธ์ (Homogeneous)

ข. การทดสอบโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น จะสามารถลดเวลาในการสอบลงได้ โดยที่ผู้สอบผ่านก็สามารถข้ามไปเรียนบทเรียนถัดไป ผู้ที่สอบไม่ผ่านก็สามารถที่จะไปศึกษาเพิ่มเติมได้ โดยไม่ต้องเสียเวลากับการทำแบบสอบที่ทราบแน่ชัดว่าไม่สามารถจะทำได้

ค. ความยากของข้อสอบเป็นปัจจัยแรกที่มีผลต่อการจัดประเภทบุคคลตามระดับการยอมรับได้อย่างถูกต้อง

ทั้งนี้ สเกลไลน์และพิชยานุวัฒน์ได้กำหนดค่า α , β , P_0 และ P_1 ขึ้นทั้งหมด 7 แผน ดังรายละเอียดในตารางที่ 10

ตารางที่ 10 แผนการกำหนดค่าพารามิเตอร์ต่างๆที่ใช้ในเทคนิคการวิเคราะห์ตามลำดับชั้น เพื่อใช้ตรวจสอบหารูปแบบของการวิเคราะห์ที่เหมาะสมที่สุดของสเกลไลน์และพิชยานุวัฒน์

แผน ค่าพารามิเตอร์	A	B	C	D	E	F	G
P_0	.05	.05	.15	.25	.25	.35	.45
P_1	.15	.35	.25	.35	.45	.45	.55
α	.05	.05	.05	.05	.05	.05	.05
β	.10	.10	.10	.10	.10	.10	.10
d_0	-2.04	-0.97	-3.54	-4.69	-2.51	-5.38	-5.61
	+.10N	+.15N	+.20N	+.30N	+.34N	+.39N	+.50N
d_1	2.62	1.24	4.54	6.03	3.22	6.91	7.19
	+.10N	+.15N	+.20N	+.30N	+.34N	+.39N	+.50N
จำนวนข้อที่น้อยที่สุด	21	7	18	16	8	14	12

จากแผนการดังกล่าวที่ได้กำหนดขึ้น สเกลไลน์และพิริยานูวัฒน์ได้เลือกใช้แผนเพื่อทดลองนำไปศึกษา ทั้งนี้เพราะว่าแผน E เป็นแผนที่ใช้ข้อสอบจำนวนน้อยที่สุดเพียง 8 ข้อ และมีข้อดีกว่าแผน B และมีค่า P_0 และ P_1 ที่ดีกว่า กล่าวคือ ค่า P_0 ไม่ต่ำจนเกินไป และค่า $P_1 - P_0$ ของแผน E มีค่าที่แคบกว่าแผน B ซึ่งมีผลทำให้ชวยังตัดสินใจไม่ได้แคบกว่าชว (Stecklein and Pitiyanuwat 1975 : 5) โดยสเกลไลน์ได้ทดลองนำไปใช้จัดประเภทระบับการรอบรู้ของนักศึกษาจำนวน 24 คน จากแบบสอบเต็มฉบับจำนวน 30 ข้อ การสอบถือเกณฑ์ว่าผู้สอบผ่านจะคงได้คะแนน 24 คะแนน หรือ 80% ขึ้นไป เมื่อนำผลที่ได้มาเทียบกับการจัดประเภทระบับการรอบรู้โดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น ปรากฏผลดังนี้

ตารางที่ 11 เปรียบเทียบผลการจัดประเภทนักศึกษาจำนวน 24 คน ความระบับการรอบรู้ระหว่างการใช้คะแนนจากแบบสอบเต็มฉบับกับผลการตัดสินใจโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้นของสเกลไลน์และพิริยานูวัฒน์

คะแนนจาก แบบสอบเต็มฉบับ	ตัดสินใจจากเทคนิคการวิเคราะห์ตามลำดับชั้น		
	รอบรู้	ยังตัดสินใจไม่ได้	ไม่รอบรู้
29	2		
28	3		
27	3		
26	1		
25	2		
24	2	1	
23	3	2	
22		1	
21		2	
20		1	
19			
18			
17			1

จากตาราง ผลการตัดสินจัดประเภทบุคคลตามระดับการรอบรู้ ปรากฏว่า มีผู้เรียนจำนวน 16 คน ที่เป็นผู้รอบรู้เมื่อตัดสินโดยใช้เทคนิคการวิเคราะห์ตามลำดับชั้น ในจำนวนนี้มีอยู่ 13 คนที่เป็นผู้รอบรู้เมื่อใช้เกณฑ์ 80% จากคะแนนจากแบบสอบเต็มฉบับ และมีผู้เข้าสอบ 1 คนที่ผลการตัดสินจากเทคนิคการวิเคราะห์ตามลำดับชั้นจัดประเภท เป็นผู้ยังตัดสินใจไม่ได้ แต่ผลการสอบจากแบบสอบเต็มฉบับจัดประเภท เป็นผู้รอบรู้ และในการทดสอบมีผู้ไม่รอบรู้ 1 คน ซึ่งได้รับการจัดประเภท เป็นผู้ไม่รอบรู้ทั้ง โดยเกณฑ์ของเทคนิคการวิเคราะห์ตามลำดับชั้น และเกณฑ์ของคะแนนจากแบบสอบเต็มฉบับ

ต่อมาในปี 1980 สเตคไลน์, พิธิยานุวัฒน์ และ ลิงก์ (Stecklein, Pitiyanuwat and Ling 1980) ได้นำเอาแผน E ที่สร้างไว้แล้วไปทดลองใช้ในสถานการณ์จริง กล่าวคือ ผู้สอบจะทราบค่าตอบของข้อสอบแต่ละข้อทันทีที่ตอบเสร็จจากกระดาษคำตอบที่ได้รับการออกแบบ โดยเฉพาะ โดยมีเงื่อนไขว่า ผู้สอบจะได้รับการตัดสินเป็นผู้รอบรู้และถือว่าสอบผ่าน ถ้าทำข้อสอบในช่วงความยาวต่างๆ เป็นไปตามเกณฑ์ใด เกณฑ์หนึ่ง ดังต่อไปนี้

- ก. ทำข้อสอบ 8 ข้อแรกถูกหมด
- ข. ทำข้อสอบ 11 ข้อแรก ถูก 10 ข้อ
- ค. ทำข้อสอบ 14 ข้อแรก ถูก 12 ข้อ
- ง. ทำข้อสอบ 17 ข้อแรก ถูก 14 ข้อ
- จ. ทำข้อสอบ 20 ข้อแรก ถูก 16 ข้อ

ทั้งนี้ ถ้าผู้สอบทำข้อสอบได้ถูกต้องตามเกณฑ์ดังกล่าวแล้ว ก็จะมีการชอว์งให้ทำแบบสอบต่อไปจนหมดทั้งฉบับ แต่ก็อาจจะยุติการสอบได้ถ้าผู้สอบต้องการ

จากการนำผลที่ได้มาศึกษาเทียบกับผลการตัดสินจากแบบสอบเต็มฉบับ พบว่า

ในรายวิชาแรก สามารถตัดสินผู้สอบได้ถูกต้องในช่วงความยาวต่างๆ ทั้ง 5 ช่วง เป็นจำนวน 68 ใน 152 คน และเกิดความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน 3 คน ในรายวิชาที่สอง จำนวนผู้เข้าสอบ 53 คน เมื่อใช้เทคนิคการวิเคราะห์ตามลำดับชั้นเป็นเกณฑ์ สามารถตัดสินให้ผู้สอบเป็นผู้รอบรู้ได้เป็นจำนวน 22 คน และเป็นผู้ไม่รอบรู้ 6 คน โดยไม่มีความคลาดเคลื่อนในการให้ผู้รอบรู้สอบตก เกิดขึ้น และเกิดความคลาดเคลื่อนในการให้ผู้ไม่รอบรู้สอบผ่าน

ร้อยละ 3.6 ส่วนในรายวิชาที่สาม สามารถคัดลึนใจให้ผูสอบสอบผ่านได้ถึง 27 ใน 31 คน
 แทอาจารย์ประจำวิชามีความเห็นว่แบบสอบในช่วงต้นๆ ประกอบขึ้นด้วยข้อสอบที่ง่ายจนเกินไป
 จากการศึกษากฎกิริยาของผูสอบที่สอบในระบบดังกล่าว ในค่านความรู้สึกที่มีต่อการ
 ทราบคำตอบทันที พบว่าผูสอบมีความเห็นตางกันในเรื่องนี้ตางๆค่อนข้างมาก กล่าวคือผูสอบ
 ร้อยละ 38 เห็นว่เป็นสิ่งที่ททาหาย ผูสอบร้อยละ 32 เห็นว่เป็นวิธีการที่น่าสนใจ แต่ในขณะที่
 เดียวกันก็มีผูสอบร้อยละ 34 เห็นว่เป็นการสอบที่เสียเวลา และผูสอบร้อยละ 32 เห็นว่เป็น
 การสอบที่สร้างภาวะความคับข้องใจ ค่านความเห็นต่อระบบการกำหนดความยาวในส่วนย่อย
 ของแบบสอบเพื่อคัดลึนผลระดับการ รณรู้ ผลที่ได้ก็มีความขัดแย้งกันมากในตัวเองเช่นกัน เช่น
 ผูสอบร้อยละ 84 เห็นว่เป็นระบบการสอบที่ทำให้เกิดความเครียด ในขณะที่ผูสอบร้อยละ 83
 เห็นว่เป็นการทดสอบที่ททาหายความสามารถ

จากงานวิจัยของส.เทคโนโลยีและพินิจาณวัฒน์ในปี 1975 และงานของส.เทคโนโลยี
 พินิจาณวัฒน์ และสิงค์ ในปี 1980 ซึ่งมีวัตถุประสงค์คือมุ่งประยุกต์เทคนิคการวิเคราะห์ตาม
 ลำดับชั้นไปใช้ใในสถานการณ์จริงในระดับชั้นเรียนนั้น สามารถกล่าวได้ว่า แนวคิดพื้นฐานตางๆ
 ที่ได้พัฒนาไว้นั้น สามารถที่จะนำมาใช้เป็นหลักการพื้นฐานในการกำหนดความยาวของแบบสอบ
 อิงเกณฑ์ใด เป็นอย่างใด ในทัศนะของผู้วิจัยเห็นว่ยังมีประเด็นที่น่าจะพิจารณาอยู่อีกประเด็นหนึ่ง
 ที่ถ้าได้มีการศึกษาและพัฒนาเพิ่มเข้าไปในระบบดังกล่าวแล้ว จะมีผลทำให้งานค้นคว้าในเรื่องนี้
 สามารถพัฒนาไปได้อีกชั้นหนึ่ง คือการคำนึงถึงค่าความคลาดเคลื่อนประเภทตางๆที่ เกิดขึ้นใ
 การจัดแบ่งผูสอบออกตามระดับการรอบรู้ระหว่างการใช้คะแนนจากแบบสอบเต็มฉบับ เป็นเกณฑ์
 กับเมื่อใช้ผลการคัดลึนจากเทคนิคการวิเคราะห์ตามลำดับชั้นเป็นเกณฑ์ โดยถ้าได้มีการวิเคราะห์
 ทำความเข้าใจ จัดประเภทความคลาดเคลื่อนตางๆที่ เกิดขึ้นให้เด่นชัด คือแยกประเภทความ
 คลาดเคลื่อนที่ปรากฏออกเป็นความคลาดเคลื่อนใในการจัดประเภทบุคคลตามระดับการรอบรู้
 ที่มีความสามารถตามจุด เกณฑ์มาตรฐาน ความคลาดเคลื่อน ๗ จุดคัดลึนใจที่กำหนด และ
 ความคลาดเคลื่อนที่แท้จริง ก็จะมีผลทำให้การประยุกต์เทคนิคการวิเคราะห์ตามลำดับชั้นเพื่อ
 กำหนดความยาวของแบบสอบอิง เกณฑ์ เป็นไปอย่างมีหลักวิชาหนักแน่นและน่าเชื่อถือยิ่งขึ้น