บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ความพยายามที่จะปรับแก้ข้อบกพร่องของข้อสอบเลือกตอบชนิดแบบฉบับ

ดังได้กล่าวมาแล้วว่าข้อสอบเลือกตอบชนิดแบบฉบับมีจุดอ่อนที่ทำให้ผลการวัดโดยใช้ข้อสอบชนิดนี้มีความถูกต้องน้อยกว่าที่ควรจะเป็นอยู่หลายประการ จุดอ่อนที่ควรได้รับการพิจารณา 3 ประการคือ

1. ปัญหาบางปัญหามีคำตอบที่ถูกต้องได้หลายคำตอบ ข้อสอบเลือกตอบชนิดแบบฉบับไม่สะดวกที่จะใช้วัดในกรณีที่เราต้องการทราบว่าผู้ตอบทราบคำตอบที่ถูกต้องทั้งหมดหรือไม่
2. การมีตัวเลือกที่ถูกต้องเพียงตัวเดียว อาจทำให้ผู้ตอบสามารถตอบข้อสอบได้ถูกต้องโดยไม่ต้องมีความรู้ในเรื่องนั้นอย่างครบถ้วนสมบูรณ์
3. การตรวจให้คะแนนไม่ละเอียดถี่ถ้วนพอที่จะแยกผู้ที่มีความรู้สมบูรณ์ครบถ้วน ผู้ที่มีความรู้บางส่วนและผู้ที่ไม่รู้ออกจากกันได้

แต่นักวัดผลก็ได้พยายามที่จะปรับปรุงแก้ไขข้อสอบเลือกตอบให้วัดได้อย่างมีประสิทธิภาพมากขึ้น ด้วยวิธีการต่าง ๆ หลายวิธี ซึ่งพอสรุปได้เป็น 3 กลุ่ม ดังนี้

1. **ปรับวิธีการตรวจให้คะแนน** วิธีการนี้จะคงรูปแบบของข้อสอบคือให้มีตัวเลือกที่เป็นคำตอบที่ถูกเพียงตัวเลือกเดียวไว้และวิธีการตอบก็ยังคงเหมือนเดิม คือ เลือกตอบเฉพาะตัวเลือกที่ถูกเพียงตัวเลือกเดียว แต่จะมาแก้ไขตรงวิธีการตรวจให้คะแนน ซึ่งก็มีวิธีการอยู่ 2 วิธี คือ

1.1 **การให้คะแนนโดยใช้สูตร** (formula scoring) สูตรสำหรับให้คะแนนมีอยู่หลายรูปแบบ แต่สูตรที่เป็นที่นิยมใช้กันมากก็คือสูตรปรับแก้การเดา $S = R - W/(N - 1)$ เมื่อ S คือคะแนนที่ได้โดยปราศจากการเดา R คือ จำนวนข้อที่ตอบถูก W คือจำนวนข้อที่ตอบผิด และ N คือ จำนวนตัวเลือกในแต่ละข้อ สูตรนี้มีข้อตกลงเบื้องต้นที่สำคัญ คือ เมื่อผู้ตอบไม่รู้คำตอบที่ถูก ผู้ตอบจะเลือกคำตอบแบบเดาสุ่ม ถ้าใช้สูตรนี้ปรับแก้แล้วคะแนนที่คาดหวัง (expected score) ในข้อนั้นจะเป็น 0 การใช้สูตรนี้ได้รับการวิจารณ์ว่าตั้งอยู่บนข้อตกลงเบื้องต้นที่ไม่สมเหตุสมผล

คือไปคิดว่าการตอบผิดนั้นเป็นผลจากการเดาสุ่ม (Ebel 1972: 250) แท้จริงแล้วการตอบผิดอาจจะเป็นเพราะ ผู้ตอบมีความรู้เรื่องนั้นเพียงบางส่วนหรือมีความรู้ที่ผิดก็ได้ ซึ่งคนเหล่านี้จะไม่เลือกคำตอบแบบเดาสุ่ม ตัวเลือกที่ผิดจะดึงดูดความสนใจของผู้ตอบเหล่านี้ไม่เท่ากัน ดังนั้นโอกาสในการตอบข้อสอบถูกหรือผิดของแต่ละคนจึงไม่น่าจะเท่ากัน จากผลการวิจัยก็พบว่า การใช้สูตรปรับแก้การเดานี้ก็ไม่ได้ทำให้ความเที่ยงหรือความตรงของผลการสอบแตกต่างไปจากการไม่ได้ปรับแก้การเดา (Lord 1968: 308-309; เพ็ญศรี สว่างเนตร 2520) แสดงว่าการใช้สูตรปรับแก้การเดาอย่างง่าย ๆ นี้ไม่ได้ช่วยแก้ปัญหาเรื่องการเดาแต่อย่างใด สูตรที่ใช้ดูจะเป็นการลงโทษผู้ที่เคราะห์ร้ายเดาแล้วผิดเสียมากกว่า (Ebel 1972: 251) สูตรปรับแก้การเดารูปแบบอื่น ๆ ที่มีผู้เสนอไว้อีก เช่น Reid (1977) ได้เสนอสูตรให้คะแนนแบบความแปรปรวนเป็นศูนย์ (zero variance scoring formula) ดังนี้

$$S = Q(n - Q/R)/(n - 1)$$

ในบทความเดียวกัน Reid ก็ได้เสนอสูตรให้คะแนนที่เป็น linear combination ของสูตรให้คะแนนแบบเดิมกับแบบความแปรปรวนเป็นศูนย์ ดังนี้

$$S = [nR + (n - 1)Q - Q^2/R]/2(n - 1)$$

เมื่อ n คือ จำนวนตัวเลือกในแต่ละข้อ

Q คือ จำนวนข้อสอบทั้งหมด

R คือ จำนวนข้อที่ตอบถูก

ส่วน Hamdan ก็ได้เสนอสูตรให้คะแนนเมื่อผู้ตอบเดาอย่างมีความรู้ (Hamdan 1979: 29-31) โดย

$$S = X_o - k/n$$

เมื่อ X_o คือ จำนวนข้อที่ตอบถูก

k คือ จำนวนข้อสอบทั้งหมด

n คือ จำนวนตัวเลือกในแต่ละข้อ

อย่างไรก็ตาม สูตรการให้คะแนน 3 สูตร สุดท้ายนี้ยังไม่พบรายงานเกี่ยวกับผลของการใช้สูตรการให้คะแนนเหล่านี้แต่อย่างใด ยังคงเป็นแต่เพียงการเสนอแนวคิดไว้เท่านั้น

1.2 **การให้น้ำหนักคะแนนของแต่ละตัวเลือกไม่เท่ากัน** วิธีการนี้จะมีการกำหนดน้ำหนักคะแนนของแต่ละตัวเลือกเอาไว้ก่อน ถ้าผู้ตอบเลือกตอบตัวเลือกใดก็จะได้คะแนนตามที่กำหนดไว้ วิธีการกำหนดน้ำหนักคะแนนอาจทำได้หลายวิธี เช่น กำหนดค่าน้ำหนักตามระดับความถูกต้องของตัวเลือก หรือกำหนดจากผลการตอบแบบสอบคู่ขนานของกลุ่มมาตรฐาน เป็นต้น จากผลการศึกษาวิธีการให้คะแนนแบบนี้รายงานว่า ทำให้การวัดมีความเที่ยงสูงขึ้น (Davis and Fiffer 1959; Patnaik and Traub 1973; Reilly and Jackson 1973) ส่วนความตรงมีค่าไม่แตกต่างไปจากการให้คะแนนแบบ 0-1 นอกจากการศึกษาของ Patnaik and Traub ที่พบว่าความตรงลดลงเล็กน้อย วิธีการแก้ปัญหาแบบนี้แม้จะมีหลักฐานว่าทำให้ความเที่ยงของการวัดสูงขึ้น แต่เมื่อคำนึงถึงการนำไปใช้แล้ว คงจะมีปัญหายุ่งยากในเรื่องของการกำหนดน้ำหนักคะแนนของแต่ละตัวเลือกให้เหมาะสม จึงไม่น่าจะเป็นวิธีการที่เหมาะสมกับการใช้วัดผลในชั้นเรียนตามปกติ

2. **ปรับวิธีการตอบและปรับวิธีการตรวจให้คะแนน** วิธีการนี้รูปแบบของข้อสอบจะยังคงเหมือนเดิมคือมีตัวเลือกที่เป็นคำตอบที่ถูกเพียงตัวเลือกเดียว แต่วิธีการตอบจะเปลี่ยนไปซึ่งวิธีการตรวจให้คะแนนก็จะปรับเปลี่ยนไปตามวิธีการตอบด้วย วิธีการนี้มีอยู่ 4 วิธีคือ

2.1 **การให้ตอบปฏิเสธตัวเลือกที่ผิด** เป็นวิธีที่เสนอโดยคูมส์กับคณะ (Coombs, Milholand, and Womer 1959) การตอบข้อสอบแบบนี้ผู้ตอบจะต้องบอกว่าตัวเลือกใดเป็นตัวเลือกที่ผิดบ้าง ส่วนตัวเลือกที่คิดว่าเป็นตัวเลือกที่ถูกหรือไม่แน่ใจให้เว้นไว้ การให้คะแนนจะให้ทุกตัวเลือกที่ทำเครื่องหมาย ถ้าตัวเลือกที่ถูกทำเครื่องหมายเป็นตัวเลือกที่ผิดก็จะได้ตัวเลือกละ 1 คะแนนตัวเลือกที่เว้นไว้จะได้ 0 คะแนน แต่ถ้าไปทำเครื่องหมายที่เป็นตัวเลือกที่ถูกก็จะเสียคะแนนเท่ากับ k - 1 (k คือ จำนวนตัวเลือก) คะแนนที่เป็นไปได้ในแต่ละข้อจะอยู่ระหว่าง 1 - k ถึง k - 1 คะแนนตามลักษณะการตอบที่แตกต่างกันซึ่งจะแสดงถึงระดับความรู้ความสามารถที่แตกต่างกัน จากการให้ตอบตอบข้อสอบและให้คะแนนโดยวิธีนี้พบว่าให้ผลการวัดที่มีความเที่ยงสูงกว่าข้อสอบเลือกตอบชนิดแบบฉบับ แต่ไม่พบว่ามีความแตกต่างกันในเรื่องของความตรง จากศึกษาของอรวรรณ ตัณฑเจริญรัตน์ (2517) และสำราญ มีแจ้ง (2525) ก็ได้ข้อค้นพบเช่นเดียวกัน แต่จากการศึกษาของกาญจนา ศิริวัฒนพงษ์ (2520) กลับพบว่าวิธีการนี้มีความเที่ยงต่ำกว่าวิธีของข้อสอบเลือกตอบชนิดแบบฉบับ แต่ในเรื่องของความตรงข้อค้นพบสอดคล้องกัน จากผลการวิจัยวิธีการของคูมส์และคณะเสนอมานี้ค่อนข้างจะมีประสิทธิภาพ ทั้งนี้อาจเนื่องจากวิธีการ

ตรวจให้คะแนนยับยั้งผู้ตอบไม่ให้เดาสุ่มค่อนข้างจะได้ผลและพิสัยของคะแนนค่อนข้างกว้าง อย่างไรก็ตามคูมส์และคณะก็ยอมรับว่าวิธีการตรวจให้คะแนนแบบนี้ออกจะยุ่งยากในการตอบ ถ้าผู้ตอบทราบคำตอบที่ถูกแต่ไม่ทราบว่าตัวเลือกบางตัวผิดเพราะอะไร เขาก็จะไม่พิจารณาตัวเลือกที่เหลือเพราะเชื่อว่าต้องเป็นคำตอบที่ผิดแน่นอน เขาก็จะสามารถทำคะแนนข้อนั้นได้เต็มเท่ากับผู้ที่ทราบทั้งคำตอบที่ถูกและคำตอบที่ผิดทั้ง ๆ ที่ผู้ตอบ 2 คนนี้ มีความรู้ไม่เท่ากัน

2.2 **การให้ตอบรับตัวเลือกที่ถูกและปฏิเสธตัวเลือกที่ผิด** เป็นวิธีการที่เสนอโดย อนันต์ ศรีโสภา (2516) วิธีการตอบผู้ตอบต้องทำเครื่องหมายเพิ่มจากวิธีการที่คูมส์และคณะเสนอ คือต้องทำเครื่องหมายถูกที่ตัวเลือกที่ถูกด้วย การให้คะแนนก็จะให้เป็นรายตัวเลือกเช่นกัน โดยตัวเลือกที่ทำเครื่องหมายถูกต้องตรงตามเฉลยจะได้ 1 คะแนน ส่วนตัวเลือกที่ทำเครื่องหมายตรงข้ามกับเฉลยก็จะได้ -1 คะแนน ตัวเลือกใดเว้นไปก็จะได้คะแนนเป็น 0 อรวรรณ ตัณฑเจริญรัตน์ (2517) และกาญจนา ศิริวัฒนพงษ์ (2520) ได้นำวิธีการให้คะแนนแบบนี้ไปศึกษาพบว่า ผลของการให้คะแนนแบบนี้มีความเที่ยงสูงกว่าวิธีการของคูมส์และคณะและสูงกว่าวิธีการให้คะแนนแบบ 0-1 ส่วนความตรงนั้นไม่แตกต่างกัน แต่จากการศึกษาของสำราญ มีแจ้ง (2525) กลับพบว่าวิธีการนี้ให้ผลการวัดที่มีความเที่ยงไม่แตกต่างจากวิธีการให้คะแนนแบบ 0-1 แต่มีความตรงสูงกว่า ส่วนค่าอำนาจจำแนกนั้นไม่แตกต่างกัน จุดอ่อนของวิธีนี้ก็เช่นเดียวกับวิธีการของคูมส์และคณะ คือถ้าผู้ตอบทราบว่าตัวเลือกใดเป็นคำตอบที่ถูกแล้ว เขาก็สามารถทำเครื่องหมายผิดที่ตัวเลือกที่เหลือและได้คะแนนเต็มในข้อนั้นได้ โดยไม่ต้องทราบว่าตัวเลือกเหล่านั้นทำไมจึงผิด จุดอ่อนอีกประการหนึ่งที่สำคัญมากของวิธีนี้คือ วิธีนี้ล่อใจให้ผู้ตอบแบบเดาค่อนข้างมาก นั่นคือ โดยทำเครื่องหมายผิดทุกตัวเลือกเขาก็จะได้คะแนนอย่างน้อย k - 2 คะแนนซึ่งเป็นคะแนนที่ค่อนข้างสูงทีเดียว วิธีนี้จึงไม่น่าจะเป็นวิธีที่จะแก้ปัญหาการใช้ข้อสอบเลือกตอบได้อย่างมีประสิทธิภาพ

2.3 **การให้เลือกคำตอบได้หลายตัวเลือก** วิธีนี้ผู้ตอบจะสามารถทำเครื่องหมายแสดงว่าตัวเลือกนั้น ๆ ถูกได้หลายตัวเลือกเท่าที่ผู้ตอบคิดว่าตัวเลือกใดควรจะเป็นคำตอบที่ถูกโดยที่คำตอบที่ถูกนั้นมีอยู่เพียงตัวเลือกเดียว การให้คะแนนนั้นคะแนนที่ได้จะลดลงไปตามจำนวนตัวเลือกที่เลือกตอบเพิ่มขึ้น ผู้ใช้วิธีนี้เชื่อว่ายิ่งผู้ตอบทำเครื่องหมายที่ตัวเลือกหลายตัวขึ้นแสดงว่าเขามีความรู้ในเรื่องนั้นน้อยลงเรื่อย ๆ ผลของการใช้วิธีการนี้ Dressel and Schmid (1953) พบว่า มีความเที่ยงและความตรงต่ำกว่าวิธีการตรวจให้คะแนนแบบ 0-1 แต่จากการศึกษาของสำราญ มีแจ้ง (2525) กลับพบว่า วิธีการนี้มีทั้งความเที่ยงและความตรงสูงกว่าวิธีการให้คะแนนแบบ 0-1 ส่วนค่าอำนาจจำแนกนั้นไม่แตกต่างกัน ผลการศึกษาที่ออกมาตรงกันข้ามกันเช่นนี้ อาจเป็นเพราะวิธีการให้คะแนนของ Dressel and Schmid ต่างจากของสำราญ มีแจ้งก็ได้

กล่าวคือ Dressel and Schmid จะให้คะแนนติดลบกับบรรดาตัวเลือกที่ทำเครื่องหมายแต่ไม่มีคำตอบที่ถูกต้องอยู่และจะติดลบมากขึ้นถ้ายิ่งเลือกหลายตัวเลือก แต่สำราญ มีแจ้ง ให้คะแนนในกรณีเช่นนี้เป็น 0 ทั้งหมด การให้คะแนนตามวิธีการของ Dressel and Schmid ดูจะไม่ค่อยสมเหตุสมผลนัก เพราะถ้าผู้ตอบทำเครื่องหมายทุกตัวเลือก (แสดงว่าไม่มั่นใจในการเลือกเลย) จะได้คะแนนมากกว่าผู้ที่เลือกตัวเลือกน้อยกว่าแต่ไม่มีคำตอบที่ถูกต้องรวมอยู่ด้วย คล้ายกับการตอบด้วยความมั่นใจกว่าแต่ตอบผิดแล้วถูกลงโทษมากกว่า วิธีการให้คะแนนของสำราญ มีแจ้ง ดูจะมีเหตุผลมากกว่า และหลักฐานเชิงประจักษ์ก็ยืนยันว่ามีประสิทธิภาพดีกว่าวิธีการให้คะแนนแบบ 0-1 ส่วนวิธีการของ Dressel and Schmid นั้นให้ผลตรงข้าม อย่างไรก็ตามข้อสอบเลือกตอบชนิดเลือกคำตอบได้ไม่จำกัดนี้ยังเป็นข้อสอบที่มีคำตอบที่มีตัวเลือกที่ถูกต้องอยู่เพียงตัวเลือกเดียว เมื่อผู้ตอบเลือกคำตอบที่ตนคิดว่าถูกได้แล้วจะไม่สนใจตัวเลือกอื่นเพราะจะถือว่าผิดทันที โดยอาจจะไม่สามารถบอกได้ว่าทำไมถึงผิด การเขียนข้อสอบที่ขาดความระมัดระวังก็อาจช่วยให้ผู้ตอบเลือกคำตอบที่ถูกต้องได้ง่ายขึ้น การตรวจให้คะแนนก็ยังถือได้ว่าไม่ง่ายนัก

2.4 **การให้ผู้ตอบบอกระดับความมั่นใจในการตอบ** วิธีการนี้เมื่อผู้ตอบเลือกคำตอบที่คิดว่าถูกแล้ว ผู้ตอบจะต้องบอกด้วยว่ารู้สึกมั่นใจแค่ไหนว่าคำตอบที่เลือกนั้นเป็นคำตอบที่ถูกต้องจริง ๆ การให้คะแนนถ้าตอบถูกต้องความมั่นใจมากที่สุดก็จะได้คะแนนสูงสุด ถ้าตอบผิดแต่มีความมั่นใจมากที่สุดจะได้คะแนนติดลบมากที่สุด แต่ถ้าตอบว่าไม่มั่นใจเลยจะได้คะแนน 0 ผลการศึกษาวิธีการให้คะแนนวิธีนี้รายงานผลตรงกันว่า ทำให้การวัดมีความเที่ยงสูงกว่าการให้คะแนนแบบ 0-1 หรือวิธีการใช้สูตรปรับแก้การเดา (Dressel and Schmid 1953; Pugh and Brunza 1975; Abu-Sayf and Diamond 1976; เน็ญศรี สว่างเนตร 2520) ส่วนเรื่องความตรงนั้น Dressel and Schmid และ Abu-Sayf and Diamond พบว่าสูงกว่าการให้คะแนนแบบ 0-1 เล็กน้อย วิธีการให้คะแนนตามระดับความมั่นใจดูจะเป็นวิธีที่น่าสนใจในการแก้ปัญหาการใช้ข้อสอบแบบเลือกตอบ แต่ก็มีข้อควรพิจารณาในการตรวจให้คะแนนแบบนี้ก็คือ การที่ผู้ตอบมีความมั่นใจในระดับสูงว่าคำตอบที่ตนเลือกเป็นคำตอบที่ถูกต้องแต่กลับปรากฏว่าผิด ผู้ตอบจะได้คะแนนน้อยกว่าผู้ที่ตอบว่าไม่มั่นใจเลยนั้นเป็นสิ่งถูกต้องและยุติธรรมหรือไม่ เพราะผู้ที่มั่นใจมากแต่ตอบผิดนั้นถือว่าเป็นผู้มีความรู้ที่ผิด (misinformation) ส่วนผู้ที่ไม่มั่นใจนั้นถือว่าตอบโดยการเดา ผู้ที่มีความรู้ที่ไม่ถูกต้องนั้นไม่น่าจะถือว่ามีความสามารถน้อยกว่าผู้ไม่รู้ ในแง่ของการนำไปใช้การให้คะแนนแบบนี้มีความยุ่งยากมากพอสมควร

3. **ปรับรูปแบบข้อสอบ วิธีการสอบ และวิธีการตรวจให้คะแนน** วิธีการนี้รูปแบบของข้อสอบจะถูกปรับให้มีตัวเลือกที่เป็นคำตอบที่ถูกได้หลายตัวเลือก หรือไม่มีเลย ส่วนวิธีการตอบและวิธีการตรวจให้คะแนนก็ปรับตามไป วิธีนี้มีอยู่ 3 วิธีคือ

3.1 **การใช้ข้อสอบเลือกตอบชนิดหลายคำตอบ** (multiple-answer) ข้อสอบแบบนี้ผู้ตอบจะต้องเลือกตัวเลือกที่เป็นคำตอบที่ถูกให้ได้ทุกตัว โดยผู้ตอบจะไม่ทราบว่าตัวเลือกที่ถูกมีกี่ตัว การให้คะแนนการตอบแบบนี้อาจทำได้หลายวิธี การตอบจะให้ทำเครื่องเฉพาะตัวเลือกที่เขาคิดว่าถูกต้องเท่านั้น จากการทดลองใช้ข้อสอบแบบนี้ Dresse and schmid (1953) พบว่า มีความเที่ยงและความตรงสูงกว่าข้อสอบชนิดคำตอบเดียว ส่วนการศึกษาของ Hsu, Moss, and Khampalikit (1984) ซึ่งได้ศึกษาถึงวิธีการตรวจให้คะแนนแบบต่าง ๆ ด้วยพบว่า ข้อสอบชนิดหลายคำตอบไม่ว่าจะตรวจให้คะแนนแบบใดก็มีความเที่ยงสูงกว่าหรืออย่างน้อยก็เท่ากับข้อสอบชนิดคำตอบเดียว ขณะเดียวกันก็มีความยากและอำนาจจำแนกสูงกว่าด้วย นอกจานนั้น Hsu และคณะยังได้เปรียบเทียบสารสนเทศที่ได้จากข้อสอบแต่ละชนิดด้วยปรากฏว่า ข้อสอบชนิดคำตอบเดียวให้สารสนเทศมากกว่าในกลุ่มผู้ตอบที่มีความสามารถต่ำกว่าปานกลาง ส่วนในกลุ่มผู้ตอบที่มีความสามารถสูงกว่าระดับปานกลางข้อสอบชนิดหลายคำตอบให้สารสนเทศมากกว่า จะเห็นว่าข้อสอบชนิดหลายคำตอบช่วยแก้ปัญหากรณีคำถามหรือตัวยืนมีคำตอบที่ถูกต้องได้มากกว่า 1 คำตอบได้ สามารถวัดได้ว่าผู้ตอบรู้คำตอบที่ถูกต้องในปัญหานั้นครบถ้วนหรือไม่ และจากข้อมูลเชิงประจักษ์ก็แสดงว่าข้อสอบชนิดนี้ค่อนข้างจะมีคุณภาพดีกว่าข้อสอบเลือกตอบชนิดแบบฉบับ แต่จุดอ่อนของข้อสอบชนิดนี้ก็คือ การถือว่าตัวเลือกที่ผู้ตอบไม่ทำเครื่องหมายที่แสดงว่าผู้ตอบเห็นว่าตัวเลือกนั้นไม่ใช่คำตอบที่ถูกต้องนั้นอาจไม่ใช่ก็ได้ การไม่ทำเครื่องหมายนั้นอาจเป็นเพราะผู้ตอบไม่ทราบว่าตัวเลือกนั้นถูกหรือผิดก็ได้ อีกประการหนึ่งก็คือตัวเลือกที่เป็นคำตอบที่ถูกต้องในแต่ละข้อมีจำนวนไม่เท่ากัน การให้คะแนนด้วยวิธีการรวมคะแนนจากตัวเลือกที่ตอบถูกจะทำให้ข้อสอบแต่ละข้อมีคะแนนไม่เท่ากัน เท่ากับเป็นการถ่วงน้ำหนักข้อสอบแต่ละข้อให้ไม่เท่ากันโดยไม่สมเหตุสมผล ถ้าจะปรับให้คะแนนแต่ละข้อเท่ากันจำเป็นต้องมีสูตรปรับแก้ เป็นการเพิ่มความยุ่งยากในการตรวจให้คะแนน อาจทำให้เกิดความไม่สะดวกในการนำไปใช้

3.2 **การใช้ข้อสอบเลือกตอบชนิดเชิงซ้อน** (complex multiple-choice) ข้อสอบเลือกตอบชนิดเชิงซ้อน เป็นการนำเอาตัวเลือกของข้อสอบชนิดหลายคำตอบมาจัดกลุ่มใหม่ได้แต่ละกลุ่มตัวเลือกนั้นมาเป็นตัวเลือกที่ผู้ตอบจะต้องเลือก โดยจะมีกลุ่มตัวเลือกที่ประกอบด้วยตัวเลือกที่เป็นคำตอบที่ถูกต้องทั้งหมดเป็นตัวเลือกที่ถูกต้องเพียงกลุ่มเดียว จากการศึกษาเปรียบเทียบการใช้ข้อสอบชนิดนี้กับข้อสอบชนิดคำตอบเดียวของ Weiten (1982) พบว่าข้อสอบชนิดเชิงซ้อนยากกว่า

ข้อสอบเลือกตอบชนิดคำตอบเดียว มีอำนาจจำแนกและความเที่ยงต่ำกว่า แต่ไม่มีความแตกต่างกันในเรื่องของความตรง ส่วน Kolstad และคณะ (1983) ได้ศึกษาโดยเปรียบเทียบกับการใช้ข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก ใน 2 รายวิชา ปรากฏว่าเมื่อพิจารณาข้อสอบที่ตอบข้อสอบชนิดเชิงซ้อนสูงกว่ากลุ่มผู้ตอบข้อสอบชนิดตัดสินคำตอบทุกตัวเลือก 2 รายวิชา Kotstad และคณะสรุปว่า ข้อสอบชนิดเชิงซ้อนให้ผลการวัดที่สูงกว่าที่ควรจะเป็น อย่างไรก็ตามผลการศึกษาของ Kolstad และคณะยังไม่ค่อยจะเชื่อมั่นได้นักเนื่องจากข้อสอบที่ใช้ศึกษามีจำนวนค่อนข้างน้อย (10 ข้อ และ 14 ข้อ) แต่ข้อสรุปของ Kostad และคณะอาจเป็นไปได้ เพราะการเอาตัวเลือกมาจัดกลุ่มใหม่นั้น ถ้าจัดกลุ่มไม่เหมาะสมแล้วอาจเป็นการชี้แนะคำตอบที่ถูกได้ ผู้ที่มีความรู้อยู่บ้างบางส่วนก็จะสามารถเลือกคำตอบที่ถูกต้องได้ง่ายขึ้น เพราะผู้ตอบทราบอยู่ว่ากลุ่มคำตอบที่ถูกต้องมีอยู่เพียงกลุ่มเดียว แต่ถ้าให้เขาเลือกคำตอบที่ถูกต้องจากคำตอบเดี่ยว ๆ ไม่ได้รวมเป็นกลุ่มให้เขาอาจเลือกไม่ได้ครบทั้งหมด จึงทำให้ผู้ที่พอมีความรู้อยู่บ้างตอบได้คะแนนสูงกว่าที่ควรจะเป็น ข้อสอบชนิดนี้จึงไม่ใช่วิธีการแก้ปัญหาของการใช้ข้อสอบแบบเลือกตอบที่ดีนัก

3.3 **การใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก** (multiple true-false) ข้อสอบเลือกตอบชนิดนี้จะมีคำตอบที่ถูกหลายคำตอบหรือไม่มีเลย ผู้ตอบจะต้องตอบว่าตัวเลือกใดบ้างเป็นตัวเลือกที่ถูกและตัวเลือกใดบ้างเป็นตัวเลือกที่ผิด จากการศึกษาของ Cronbach (1941) ได้เปรียบเทียบผลการใช้ข้อสอบชนิดนี้กับข้อสอบเลือกตอบชนิดหลายคำตอบ ไม่พบว่ามีความแตกต่างกันแต่อย่างใด ไม่ว่าจะในเรื่องของเวลาที่ใช้ในการตอบ ความยากของแบบสอบ ความตรง หรือความเที่ยง ส่วน Kolstad และคณะ (1983) ได้เปรียบเทียบการใช้ข้อสอบชนิดนี้กับข้อสอบเลือกตอบชนิดเชิงซ้อน ปรากฏว่าคะแนนเฉลี่ยของข้อสอบชนิดต่ำกว่าข้อสอบชนิดเชิงซ้อน สำหรับผู้ที่ศึกษาผลการใช้ข้อสอบชนิดนี้เปรียบเทียบกับข้อสอบเลือกตอบชนิดแบบฉบับ ก็มี Frisbie and Sweeney พบว่า ผู้ตอบข้อสอบชนิดนี้ใช้เวลามากกว่าในการตอบข้อสอบชนิดแบบฉบับเล็กน้อย ผลการสอบมีความเที่ยงสูงกว่า แต่ความยากนั้นไม่อาจสรุปได้ว่าข้อสอบชนิดแบบใดยากกว่ากัน ส่วนในประเทศไทยได้มีผู้ศึกษาการใช้ข้อสอบชนิดนี้คือ จินดา โตอนันต์ (2526) พบว่า ข้อสอบชนิดนี้มีอำนาจจำแนกต่ำกว่าข้อสอบเลือกตอบชนิดแบบฉบับส่วนความเที่ยงนั้นไม่อาจสรุปผลได้ และนักเรียนที่มีความสามารถในระดับต่ำชอบและอยากให้ครูใช้ข้อสอบชนิดนี้มากกว่านักเรียนที่มีความสามารถระดับปานกลางและสูง

จะเห็นว่าวิธีการแก้ปัญหาของการใช้ข้อสอบแบบเลือกตอบแต่ละวิธีก็มีจุดเด่นจุดด้อยแตกต่างกันไป วิธีที่ผู้วิจัยเห็นว่าจะแก้ปัญหาทั้ง 3 ประการของข้อสอบเลือกตอบชนิดแบบฉบับได้โดยไม่

ยุ่งยากนักก็คือการใช้ข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือก และจากผลการศึกษาเชิงประจักษ์เกี่ยวกับประสิทธิภาพของข้อสอบแบบนี้ก็ยังไม่สามารถสรุปได้แน่ชัด จึงเห็นว่าควรจะได้มีการศึกษาเกี่ยวกับประสิทธิภาพของข้อสอบชนิดนี้โดยอาศัยกรอบทฤษฎีเหมาะสมกว่า

ทฤษฎีการตอบสนองต่อข้อสอบ

ทฤษฎีการตอบสนองต่อข้อสอบ (item response theory) หรือทฤษฎีความสามารถแฝง (latent trait theory) หรือทฤษฎีเส้นโค้งลักษณะของข้อสอบ (item characteristic curve theory) เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ภายในตัวบุคคล (Lord and Novick 1968: 358) และพยายามพัฒนาแบบจำลองเพื่อกำหนดลักษณะเฉพาะของความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบซึ่งเป็นสิ่งสังเกตได้โดยตรงกับลักษณะหรือความสามารถภายในตัวบุคคล ซึ่งไม่สามารถจะสังเกตได้โดยตรง แบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าวเป็นแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (Y) แทนพฤติกรรมการตอบสนองต่อข้อสอบ และ θ แทนลักษณะหรือความสามารถภายในตัวบุคคล ฟังก์ชันการถดถอยของ Y บน θ จะเป็นฟังก์ชันทางคณิตศาสตร์ที่จะแสดงว่าพฤติกรรมการตอบสนองต่อข้อสอบขึ้นอยู่กับระดับความสามารถอย่างไร ฟังก์ชันนี้เรียกว่า ฟังก์ชันการตอบสนองต่อข้อสอบ (item response function) (Lord 1980: 12) หรือ ฟังก์ชันลักษณะของข้อสอบ (item characteristic function) (Lord and Novick 1968: 360) ซึ่งจะกำหนดได้หลายรูปแบบขึ้นอยู่กับข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลจากการทดสอบ (Hambleton and Cook 1977: 75)

มโนทัศน์ที่สำคัญของทฤษฎีการตอบสนองต่อข้อสอบ

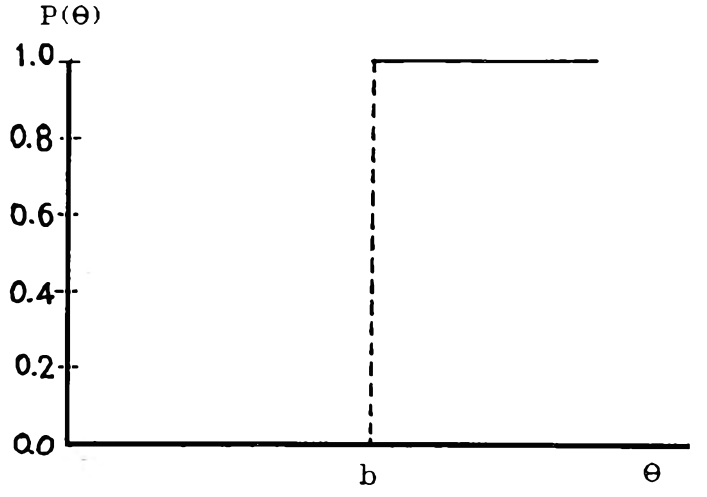
1. **มิติของลาเทนท์สเปส** (dimensionality of latent space) ลักษณะหรือความสามารถที่จะเป็นตัวกำหนดพฤติกรรมการตอบสนองต่อข้อสอบข้อใดข้อหนึ่งอาจมีได้หลายลักษณะ ลักษณะทั้งหมดเหล่านี้รวมเรียกว่า **ลาเทนท์สเปส** (latent space) จำนวนลักษณะทั้งหมดในลาเทนท์สเปสก็คือ มิติของลาเทนท์สเปส ในทฤษฎีการตอบสนองต่อข้อสอบโดยทั่วไปมักจะถือว่าลักษณะหรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบสนองต่อข้อสอบมีเพียงลักษณะเดียว ยอมรับกันเป็นข้อตกลงเบื้องต้นที่สำคัญอันหนึ่งของแบบจำลองหลายแบบในทฤษฎีนี้ เรียกข้อตกลงเบื้องต้น

นี้ว่า Unidimensionality Assumption การกำหนดข้อตกลงเช่นนี้อาจจะขัดแย้งกับความจริงที่เราทราบกันดีว่า ผลการตอบข้อสอบนั้นขึ้นอยู่กับปัจจัยหลายอย่างประกอบกัน แต่ข้อตกลงนี้ต้องการแต่เพียงว่ามีความสามารถหรือลักษณะที่เด่นและสำคัญ (dominant) เพียงลักษณะเดียวที่เป็นตัวกำหนดพฤติกรรมการตอบสนองต่อข้อสอบ ลักษณะอื่น ๆ ที่อาจจะเป็นตัวกำหนดพฤติกรรมการตอบข้อสอบนั้นถือว่าเป็นลักษณะที่ไม่สำคัญ การถือว่าลาเทนท์สเปสที่สนใจมีเพียงมิติเดียวจะช่วยแบบจำลองของทฤษฎีนี้มีความซับซ้อนน้อยลงและง่ายแก่การแปลความหมายของคะแนนจากแบบสอบ อย่างไรก็ตามก็มีผู้พยายามที่จะพัฒนาแบบจำลองที่ไม่ต้องมีข้อตกลงเบื้องต้นที่ว่า ลาเทนสเปสมีเพียงมิติเดียว เรียกว่า Multidimensional Item Response Models แต่การพัฒนาแบบจำลองเหล่านี้ยังไม่ถึงระดับที่จะนำไปประยุกต์ใช้ในทางปฏิบัติได้ (Gruijter, and Kamp 1984: 160)

2. **ความเป็นอิสระในการตอบสนองต่อข้อสอบ** (local independence) หมายความว่าพฤติกรรมการตอบสนองต่อข้อสอบข้อต่าง ๆ ในแบบสอบของบุคคลใดบุคคลหนึ่งมีความเป็นอิสระในเชิงสถิติ คือการตอบสนองต่อข้อสอบข้อหนึ่งข้อใดของผู้ตอบคนหนึ่งไม่มีผลต่อการตอบสนองต่อข้อสอบข้ออื่น ๆ ในแบบสอบ ถือเป็นข้อตกลงเบื้องต้นอีกประการหนึ่งของทฤษฎีการตอบสนองต่อข้อสอบสำหรับทุกแบบจำลอง เป็นข้อตกลงเบื้องต้นที่เป็นประโยชน์ต่อการประมาณค่าพารามิเตอร์ของข้อสอบและการประมาณค่าความสามารถของผู้สอบ

3. **โค้งลักษณะของข้อสอบ** (item characteristic curve) เป็นกราฟของฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถหรือลักษณะที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะของข้อสอบมีได้หลายรูปแบบขึ้นอยู่กับความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสที่จะตอบถูก ทำให้มีแบบจำลอง (Model) ที่ใช้อธิบายความสัมพันธ์ดังกล่าวในทฤษฎีการตอบสนองต่อข้อสอบหลายแบบจำลองด้วยกัน เช่น

Guttman Perfect Scale เป็นแบบจำลองของทฤษฎีการตอบสนองต่อข้อสอบในยุคแรก ๆ แบบจำลองนี้เชื่อว่า โอกาสในการตอบข้อสอบข้อหนึ่งข้อใดถูกมีความสัมพันธ์กับความสามารถของบุคคลในลักษณะของฟังก์ชันแบบขั้นบันได (step function) เส้นโค้งลักษณะของข้อสอบจะมีลักษณะดังภาพที่ 1



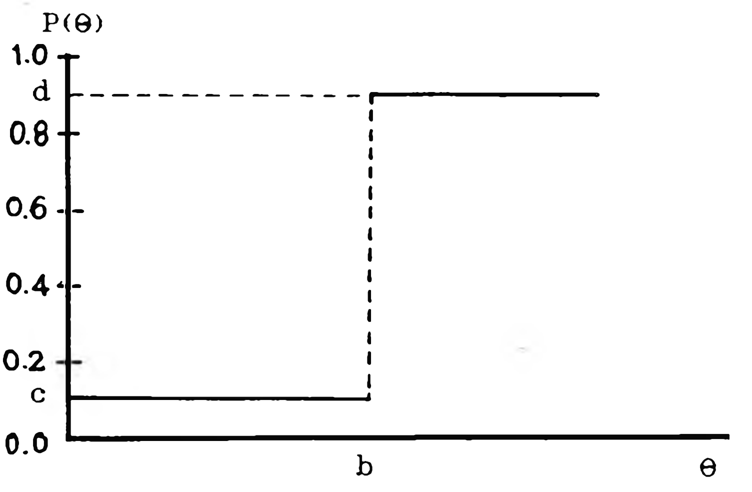
ภาพที่ 1 เส้นโค้งลักษณะข้อสอบที่เป็น Guttman Perfect Scale

จากภาพที่ 1 หมายความว่า ผู้ที่มีความสามารถต่ำกว่า b ($\theta < b$) โอกาสที่จะตอบข้อนี้ถูกจะเท่ากับ 0 ส่วนผู้ที่มีความสามารถเท่ากับหรือมากกว่า b ($\theta \geq b$) โอกาสที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 1 แบบจำลองนี้อาจเขียนสมการแสดงความสัมพันธ์ระหว่าง Θ กับโอกาสที่จะตอบข้อสอบถูก (P(Θ)) ได้ดังนี้

(1) $$P(\theta) = \begin{cases} 1 & \text{ถ้า } \theta \geq b \\ 0 & \text{ถ้า } \theta < b \end{cases}$$

แบบจำลอง Guttman Perfect Scale นี้ มีลักษณะเป็น deterministic model เป็นไปได้ยากมากที่ข้อมูลจากการสอบจะสอดคล้องกับแบบจำลอง แบบจำลองนี้จึงมีการนำไปประยุกต์ในการปฏิบัติค่อนข้างน้อย

Latent Distance Model มีลักษณะเป็น stochastic model ความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับความสามารถยังคงมีลักษณะเป็นฟังก์ชันแบบขั้นบันได แต่โอกาสในการตอบข้อสอบถูกเมื่อผู้ตอบมีความสามารถมากกว่า b จะน้อยกว่า 1 และโอกาสในการตอบข้อสอบถูกเมื่อผู้ตอบมีความสามารถน้อยกว่า b จะมากกว่า 0 เส้นโค้งลักษณะของข้อสอบจึงมีลักษณะดังแสดงในภาพที่ 2



ภาพที่ 2 เส้นโค้งลักษณะของข้อสอบใน Latent Distance Model

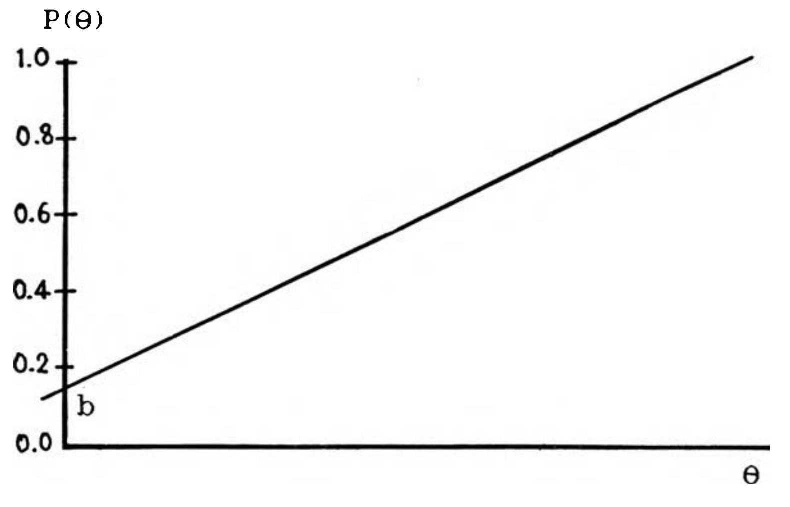
จากภาพที่ 2 แสดงว่า ผู้ที่มีความสามารถต่ำกว่า b จะมีโอกาสตอบข้อสอบข้อนี้ถูกเท่ากับ c ส่วนผู้ที่มีความสามารถเท่ากับหรือมากกว่า b จะมีโอกาสตอบข้อสอบข้อนี้ถูกเท่ากับ d โดยที่ $0 < c < d < 1$ ความสัมพันธ์ระหว่างโอกาสตอบถูกกับความสามารถเขียนเป็นสมการได้ดังนี้

(2) $$P(\theta) = \begin{cases} d & \text{ถ้า } \theta \geq b \\ c & \text{ถ้า } \theta < b \end{cases}$$

แม้ว่าแบบจำลองนี้จะเป็น stochastic model แต่ก็เป็นไปได้ยากที่ข้อมูลจริงจากการสอบจะสอดคล้องกับกับจำลองนี้ เพราะว่าบุคคลที่มีความสามารถมากกว่าหรือน้อยกว่า b จะแสดงพฤติกรรมการตอบข้อสอบที่คงเส้นคงวาเช่นนี้คงเป็นไปได้ยากมาก

Linear Model เป็นแบบจำลองที่เชื่อว่าเส้นโค้งลักษณะของข้อสอบมีลักษณะเป็นเส้นตรง ดังภาพที่ 3 โดยความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถเขียนเป็นสมการได้ดังนี้

(3) $$P(\theta) = b + a\theta$$



ภาพที่ 3 เส้นโค้งลักษณะของข้อสอบของ Linear Model

จากสมการ P(Θ) = b + a Θ ถ้า a ≠ 0 แล้ว ผู้ที่มีความสามารถในระดับต่ำมาก ๆ อาจจะมีโอกาสตอบข้อสอบถูกมีค่าเป็นลบและผู้ที่มีความสามารถในระดับสูงมากก็อาจจะมีโอกาสตอบข้อสอบถูกมีค่ามากกว่า 1 ได้ ตามนิยามค่าโอกาส (probability) จะมีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น Lazarsfeld ผู้ที่เสนอแบบจำลองนี้ก็ได้ตระหนักถึงเรื่องนี้จึงได้เสนอข้อตกลงเบื้องต้นของแบบจำลองนี้เพิ่มขึ้นอีกว่า การแจกแจงของความสามารถ (Θ) อยู่ในช่วงที่ทำให้ P(Θ) มีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น (Togerson 1958: 368) ทำให้แบบจำลองนี้มีข้อจำกัดในการนำไปใช้มาก

แบบจำลองทั้ง 3 ที่กล่าวมาแล้ว เป็นแบบจำลองในระยะแรก ๆ ของการพัฒนาทฤษฎีการตอบสนองต่อข้อสอบ แม้ว่าจะเป็นแบบจำลองที่ไม่ค่อยจะมีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากนัก แต่ก็เป็นแบบจำลองที่ให้แนวคิดเชิงทฤษฎีที่มีคุณค่าต่อการพัฒนาแบบจำลองที่เหมาะสมในระยะต่อมา

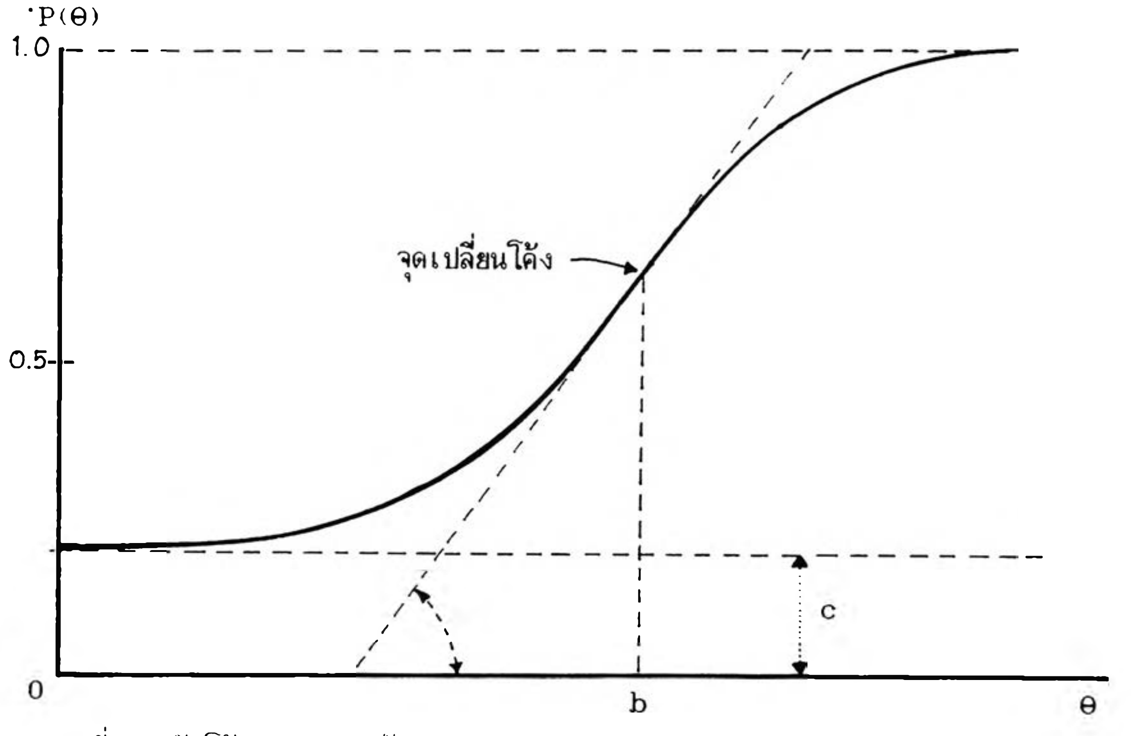
Normal Ogive Model แบบจำลองนี้เชื่อว่าความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถอยู่ในรูปของฟังก์ชันการแจกแจงสะสมปกติ (cumulative normal ogive) ซึ่งแสดงได้ด้วยสมการ

$$(4) \qquad P_i(\Theta) = \Phi[a_i(\Theta - b_i)] = \int_{-\infty}^{a_i(\Theta-b_i)} (2\pi)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}t^2]dt$$

เมื่อ a_i และ b_i เป็นค่าพารามิเตอร์ที่แสดงถึงลักษณะของข้อสอบ เส้นโค้งลักษณะของข้อสอบจะมีลักษณะคล้ายรูปตัว S โดยที่ค่าโอกาสของการตอบข้อสอบถูกจะเข้าใกล้ 0 เมื่อ θ มีค่าน้อย ๆ และจะมีค่าเพิ่มขึ้นเรื่อย ๆ และจะเข้าใกล้ 1 เมื่อ θ มีค่ามาก ๆ ในสมการ (4) ค่าพารามิเตอร์ของข้อสอบมีเพียง 2 ตัว คือ a_i กับ bi จึงเรียกแบบจำลองที่อธิบายด้วยสมการนี้ว่า Two-Parameter Normal Ogive Model แต่ในการทดสอบที่ใช้ข้อสอบแบบปรนัยที่ผู้สอบที่มีความสามารถน้อย ๆ ก็มีโอกาสตอบข้อสอบถูกโดยการเดา แบบจำลองแบบพารามิเตอร์ 2 ตัว ไม่เหมาะที่จะใช้อธิบายสถานการณ์การสอบเช่นนี้ เพราะในกรณีนี้เมื่อ θ มีค่าน้อย ๆ ค่า $P_i(\theta)$ จะไม่เข้าใกล้ 0 อีกต่อไป แต่จะมีค่าเข้าใกล้ค่า ๆ หนึ่งที่มากกว่า 0 จึงได้มีการเพิ่มค่าพารามิเตอร์ของข้อสอบเข้าไปอีก 1 ตัว คือ c_i และสมการของแบบจำลองจึงเป็น

$$(5) \qquad P_i(\theta) = c_i + (1 - c_i)\int_{-\infty}^{a_i(\theta - b_i)} (2\pi)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}t^2]dt$$

แบบจำลองนี้เรียกกว่า Three-Parameter Normal Ogive Model และเส้นโค้งลักษณะของข้อสอบก็จะมีลักษณะดังภาพที่ 4



ภาพที่ 4 เส้นโค้งลักษณะของข้อสอบของ 3-parameter normal ogive model

ค่าพารามิเตอร์ของข้อสอบจะเป็นดัชนีบอกถึงลักษณะของข้อสอบและทำให้รูปร่างของโค้งลักษณะของข้อสอบแต่ละข้อแตกต่างกันไป ค่าพารามิเตอร์แต่ละตัวมีความหมายดังนี้

a_i มีค่าเป็นสัดส่วนโดยตรงกับความชันของเส้นโค้งลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง (ความชัน ณ จุดเปลี่ยนโค้ง = $a_i(1 - c_i)/\sqrt{2\pi}$) จะเป็นค่าพารามิเตอร์ที่บ่งชี้ถึงคุณภาพของข้อสอบในแง่ของสารสนเทศที่จะได้ในการประมาณค่าระดับความสามารถ (θ) จึงอาจเรียกได้ว่าเป็น **ค่าอำนาจจำแนกของข้อสอบ** มีค่าที่เป็นไปได้อยู่ระหว่าง 0 ถึง + ∞ ในทางปฏิบัติ ค่า a ของข้อสอบโดยทั่วไปจะมีค่าระหว่าง 0.50 ถึง 2.50(Warm 1978: 52)

b_i เป็นค่าของระดับ θ ระดับหนึ่งที่ทำให้เส้นโค้งลักษณะของข้อสอบมีการเปลี่ยนโค้ง [จุดเปลี่ยนโค้งจะอยู่ที่ (b,P(b))] จะเป็นค่าพารามิเตอร์ที่บอกตำแหน่งของเส้นโค้งลักษณะของข้อสอบ ณ θ = b โอกาสที่จะตอบข้อสอบถูกจะมีค่า $(1- c_i)/2$ หรือเท่ากับ 0.5 ถ้า c_i = 0 จึงอาจเรียกได้ว่าเป็น **ค่าความยากของข้อสอบ** มีค่าที่เป็นไปได้ตั้งแต่ - ∞ ถึง + ∞ ในทางปฏิบัติมักจะพิจารณากันในช่วง -3 ถึง +3

· c_i เป็นค่ากำกับเส้นโค้งที่ต่ำสุด (lower assymtote) ของโค้งลักษณะของข้อสอบ แสดงถึงค่าโอกาสที่บุคคลที่ไม่มีความสามารถเลย ($\theta = -\infty$) จะตอบข้อสอบถูก จึงอาจเรียกได้ว่าเป็น **ค่าการเดา** (guessing parameter) หรือ pseudo-chance score level (Lord 1980: 12) มีค่าที่เป็นไปได้อยู่ระหว่าง 0 ถึง 1 ถ้าข้อสอบข้อนั้นไม่สามารถที่จะตอบถูกได้ด้วยการเดาแล้ว c_i จะเท่ากับ 0

Logistic Model เส้นโค้งลักษณะของข้อสอบมีลักษณะเป็นรูปตัว S เช่นเดียวกับ Normal Ogive Model แต่ใน Logistic Model ความสัมพันธ์ระหว่างโอกาสในการตอบข้อข้อสอบถูกกับระดับความสามารถอยู่ในรูปของฟังก์ชันการแจกแจงสะสมแบบโลจิสต์ (logistic cumulation distribution function) แสดงได้ด้วยสมการดังนี้

(6) $$P_i(\theta) = c_i + (1 - c_i)\,[\,1 + e^{-1.7a_i(\theta - b_i)}\,]^{-1}$$ (Lord 1980: 12)

โดย a_i , b_i และ c_i เป็นค่าพารามิเตอร์ของข้อสอบมีความหมายเช่นเดียวกับ Normal

Ogive Model ส่วน e คือ ค่าคงที่มีค่าประมาณ 2.71828... เส้นโค้งลักษณะของข้อสอบของ Normal Ogive Model กับของ Logistic Model จะมีลักษณะใกล้เคียงกันมาก เนื่องจากค่าฟังก์ชัน normal ogive กับค่าฟังก์ชัน logist จะแตกต่างกันเล็กน้อยเมื่อมีการปรับค่าตัวแปรด้วย scaling factor (มีค่า = 1.7) นั่นคือ $|\Phi(x) - \Psi(1.7x)| < 0.01$ สำหรับทุกค่าของ x และเมื่อ Φ แทน normal ogive function และ Ψ แทน logistic function ในแง่ของการคำนวณแล้ว Logistic Model มีความง่ายและสะดวกกว่ามาก นอกจากนี้แล้วในสถานการณ์สอบจริงอาจจะมีผู้ที่มีความสามารถสูงตอบข้อสอบผิดด้วยความเลินเล่อ กรณีเช่นนี้ Logistic Model มีความแกร่งต่อข้อมูลแบบนี้มากกว่า Normal Ogive Model จึงทำให้ Logistic Model เป็นที่นิยมกันมากในการปฏิบัติงานจริง (Lord 1980: 14)

สมการ (6) เรียกว่าเป็น Three-Parameter Logistic Model เพราะมีพารามิเตอร์ของข้อสอบ 3 ตัว ในสถานการณ์การสอบที่ผู้ตอบมีโอกาสตอบถูกด้วยการเดาน้อยมาก อาจกำหนดให้ค่าพารามิเตอร์การเดา (c_i) เป็น 0 สมการก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ a_i และ b_i สมการที่ (6) ก็จะเขียนได้ใหม่เป็น

(7) $$P_i(\Theta) = 1/[\,1 + e^{-1.7a_i(\Theta - b_i)}\,]$$

สมการนี้เรียกว่า Two-Parameter Logistic Model

ในกรณีที่กำหนดให้ $c_i = 0$ และถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน ($a_i = \bar{a}$) สมการก็จะเหลือค่าพารามิเตอร์เพียง 1 ตัวคือ b_i สมการที่ (6) ก็จะเขียนได้เป็น

(7) $$P_i(\Theta) = 1/[\,1 + e^{-1.7\bar{a}(\Theta - b_i)}\,]$$

สมการนี้เรียกว่า One-parameter Logistic Model

Rasch Model เป็นแบบจำลองที่ได้รับความนิยมมากแบบจำลองหนึ่ง พัฒนาขึ้นมาโดย Gorge Rasch นักคณิตศาสตร์ชาวเดนมาร์ก ที่ได้เสนอแบบจำลองในรูปของสมการ

(9) $P_i(\theta^*) = \theta^* / (\theta^* + b_i^*)$ (Hambleton and Cook 1977: 82)

โดย $\theta^* = e^{1.7\bar{a}\theta}$ และ $b_i^* = e^{1.7\bar{a}b_i}$ เมื่อพิจารณาแล้วจะเห็นว่า สมการ (9) ของ Rasch Model ก็คือ สมการที่ (6) ของ One-parameter Logistic Model นั่นเอง

Nominal Response Model แบบจำลองที่เสนอมาก่อนหน้านี้เป็นแบบจำลองที่ใช้กับข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (binary item) คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน Nominal Response Model เป็นแบบจำลองที่ประยุกต์ใช้กับข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า (multichotomously scored) จุดหมายของแบบจำลองนี้ก็เพื่อเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้เข้าสอบโดยใช้ประโยชน์จากสารสนเทศทั้งจากการตอบข้อสอบถูกหรือผิด แบบจำลองนี้แต่ละตัวเลือกจะมีโค้งลักษณะของตัวเลือก (item-option characteristic curve) โค้งของตัวเลือกที่เป็นคำตอบที่ถูกจะมีลักษณะเป็นฟังก์ชันที่มีค่าเพิ่มขึ้นโดยตลอด (monotonic increasing function) ส่วนโค้งของตัวลวงจะมีลักษณะอย่างไรขึ้นอยู่กับการเลือกตัวเลือกนั้นของผู้เข้าสอบที่มีความสามารถในระดับต่าง ๆ รูปแบบทางคณิตศาสตร์ของโค้งลักษณะของตัวเลือกมีได้หลายแบบ ตัวอย่างเช่น รูปแบบของ Bock ที่สมมุติว่าโอกาสที่ผู้ที่มีความสามารถเท่ากับ θ จะเลือกตอบตัวเลือก k (จากตัวเลือกทั้งหมด m ตัว) ของข้อสอบข้อที่ i กำหนดโดยสมการ

$$(10) \qquad P_{ik}(\theta) = \frac{e^{b_{ik}^* + a_{ik}^*\theta}}{\sum_{k=1}^{m} e^{b_{ik}^* + a_{ik}^*\theta}} \qquad \text{(Hambleton, et al. 1978: 476)}$$

เมื่อ b^*_{ik} และ a^*_{ik} คือค่าพารามิเตอร์ของตัวเลือก k สำหรับระดับความสามารถ θ ใด ๆ ผลรวมของโอกาสในการเลือกตัวเลือกทุกตัวจะเท่ากับ 1 กรณีที่ข้อสอบมี 2 ตัวเลือก (m = 2) แบบจำลองนี้ก็คือ Two-Parameter Logistic Model นั่นเอง

Grade Response Model เป็นแบบจำลองที่ใช้กับสถานการณ์การสอบที่พฤติกรรมตอบสนองสามารถจัดประเภทเรียงตามอันดับได้ Samejima ได้เสนอแบบจำลองที่สมมุติว่า พฤติกรรมการตอบสนองต่อข้อสอบสามารถจัดแบ่งออกได้เป็น $m_i + 1$ ประเภท และให้คะแนน $x_i = 0, 1, \ldots, m_i$ ตามลำดับ โอกาสที่ผู้มีความสามารถระดับ θ จะได้คะแนนข้อนี้เป็น x_i คือ

(11) $P_x(\Theta) = P_x^*(\Theta) + P^*_{(x_i+1)}(\Theta)$ (Hambleton, et al. 1978: 476)

เมื่อ $P_x^*(\Theta)$ คือ item response function ของการให้คะแนนแบบ 2 ค่า โดยผู้ที่ได้คะแนนต่ำกว่า x_i ถือเป็น 0 และผู้ที่ได้คะแนนเท่ากับ x_i หรือมากกว่าถือเป็น 1 ส่วนรูปแบบของฟังก์ชัน $P_x^*(\Theta)$ จะใช้แบบจำลองที่ใช้กับข้อสอบที่ให้คะแนน 2 ค่าแบบจำลองใดขึ้นอยู่กับผู้ใช้จะเลือก

การประมาณค่าพารามิเตอร์ของแบบจำลอง

Hambleton และคณะ (1977: 477-487) ได้สรุปวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองในทฤษฎีการตอบสนองต่อข้อสอบ ว่ามีวิธีใหญ่ๆ 3 วิธี คือ

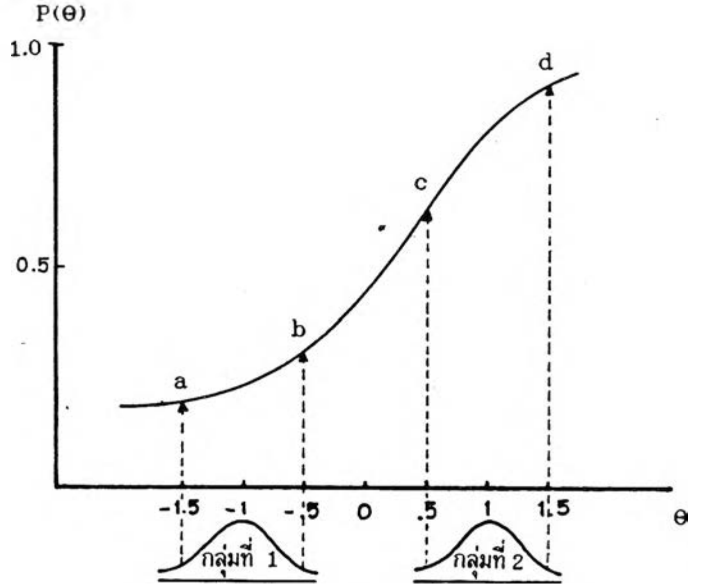
1. วิธีแมกซิมัมไลลิฮูด (Maximum Likelihood Estimation)
2. วิธีฮิวริสติค (Heuristic Estimation)
3. วิธีของเบย์ (Baysian Estimation)

วิธีการที่นิยมใช้กันมากเพราะมีประสิทธิภาพค่อนข้างสูงก็คือ วิธีแมกซิมั่มไลลิฮูด ได้มีการพัฒนาโปรแกรมคำสั่งสำเร็จรูปเพื่อใช้คอมพิวเตอร์ประมาณค่าพารามิเตอร์ด้วยวิธีนี้ขึ้นมาใช้กันหลายโปรแกรม เช่น LOGIST, BICAL, DATAGEN (Hambleton, and Cook 1977: 89-90) และ NORMOG (Hulin, Drasgow, and Parsons 1983: 53) เป็นต้น

ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ในแบบจำลอง

ประเด็นหนึ่งที่มักจะยกมาอ้างกันเสมอเมื่อกล่าวถึงความเหนือกว่า (superiority) ของทฤษฎีการตอบสนองต่อข้อสอบเมื่อเทียบกับทฤษฎีการสอบแบบคลาสสิคก็คือ ความไม่แปรเปลี่ยน (invariant) ของค่าพารามิเตอร์ของแบบจำลองไม่ว่าจะประมาณค่ามาจากกลุ่มตัวอย่างผู้ตอบข้อสอบที่มีความสามารถระดับใดก็ตาม ปรากฏการณ์นี้อธิบายได้ดังนี้คือ ฟังก์ชันการตอบสนองต่อข้อสอบ (item response function) ซึ่งเป็นฟังก์ชันที่ใช้อธิบายความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบกับระดับความสามารถนั้นเป็นฟังก์ชันการถดถอย (regression function) ของคะแนนจากข้อสอบบนคะแนนความสามารถ ตามทฤษฎีสถิติฟังก์ชันการถดถอยจะ

ไม่เปลี่ยนแปลงไปตามการแจกแจงความถี่ของตัวแปรทำนาย ดังนั้นค่าพารามิเตอร์ต่าง ๆ ในสมการถดถอยซึ่งจะเป็นตัวกำหนดรูปร่างลักษณะของเส้นถดถอยจึงไม่แปรเปลี่ยนด้วย (Lord 1980: 34) นั่นคือ ค่าพารามิเตอร์ของข้อสอบซึ่งเป็นพารามิเตอร์กำหนดลักษณะของเส้นโค้งลักษณะของข้อสอบจะไม่แปรเปลี่ยนไปไม่ว่าจะประมาณค่าจากกลุ่มผู้ตอบกลุ่มใดก็ตาม มโนทัศน์ในเรื่องนี้อธิบายได้ด้วยภาพที่ 5



ภาพที่ 5 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มตัวอย่างที่มีระดับความสามารถต่างกัน (จาก Baker 1977: 170)

ในกรณีที่ประมาณค่าจากกลุ่มที่ 1 ซึ่งเป็นกลุ่มที่มีความสามารถต่ำ เส้นโค้งลักษณะข้อสอบที่ได้จะอยู่ในช่วง a - b ซึ่งจะตกอยู่ในช่วงพิสัยของความสามารถของกลุ่มที่ 1 เส้นโค้งลักษณะข้อสอบในช่วงนี้จะบอกถึงโอกาสที่จะตอบข้อสอบข้อนี้ถูกในช่วงความสามารถที่เป็นพิสัยของกลุ่มที่ 1 ส่วนเส้นโค้งลักษณะข้อสอบในช่วง c - d ก็จะแสดงถึงโอกาสที่จะตอบข้อสอบข้อนี้ถูกในช่วงความสามารถที่เป็นพิสัยของกลุ่มที่ 2 ซึ่งเป็นกลุ่มที่มีความสามารถสูง ถ้าข้อสอบข้อนี้วัดความสามารถอันเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว เส้นโค้งลักษณะข้อสอบย่อมจะต้องมีเพียงเส้นเดียว เส้นโค้งในช่วง a - b และช่วง c - d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อสอบของข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ที่กำหนดรูปร่างลักษณะของเส้นโค้งนี้ย่อมมีค่าได้ตัวละ 1 ค่าเท่านั้น นั่นคือการประมาณค่าพารามิเตอร์ของข้อสอบที่ได้จากกลุ่มตัวอย่างทั้งสองควรจะเป็นค่าเดียวกัน

อย่างไรก็ตามการไม่แปรเปลี่ยนของค่าพารามิเตอร์นี้ไม่ได้หมายความว่า ค่าพารามิเตอร์ที่ได้จากการประมาณค่าโดยใช้กลุ่มตัวอย่างที่ต่างกันจะมีค่าเท่ากันพอดี ค่าพารามิเตอร์ที่ประมาณค่าจากการใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากัน ก็ต่อเมื่อเราเลือกมาตรวัดความสามารถที่มีจุดเริ่มต้นที่เดียวกันและหน่วยเดียวกันเท่านั้น ถ้ามาตรวัดความสามารถมีจุดเริ่มต้นและหน่วยการวัดที่ต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์จะหมายความว่า ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกันของข้อสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง (Lord 1980: 36)

ค่าสารสนเทศ

ในสถานการณ์ทั่วไป ถ้ามั่นใจว่าเหตุการณ์อันหนึ่งจะเกิดขึ้นค่อนข้างมากแสดงว่ามีข่าวสารข้อมูลหรือสารสนเทศเกี่ยวกับเหตุการณ์นั้นมากพอสมควร ในทางกลับกันถ้าไม่มีข่าวสารข้อมูลเกี่ยวกับเหตุการณ์นั้นหรือมีน้อยความมั่นใจก็จะมีน้อยตามไปด้วย ในการอ้างอิงเชิงสถิติความแม่นยำของการประมาณค่าพารามิเตอร์ของกลุ่มประชากรอาจดูได้จากช่วงกว้างของค่าประมาณ ถ้าไม่มีสารสนเทศใด ๆ เกี่ยวกับประชากรเลยก็อาจจะต้องประมาณค่าเป็นค่าใด ๆ ในช่วง $-\infty$ ถึง $+\infty$ แต่ถ้ามีสารสนเทศเกี่ยวกับประชากรบ้าง ช่วงของค่าประมาณจะแคบเข้า นั่นคือความแม่นยำในการประมาณเริ่มมีมากขึ้น ตามปกติความแม่นยำของการประมาณค่าพารามิเตอร์จะแสดงได้ด้วยค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า คือ ถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีมาก ความแม่นยำของการประมาณค่าก็จะมีน้อย เพราะช่วงของค่าประมาณจะกว้าง ในทางกลับกันถ้าความคลาดเคลื่อนมาตรฐานของการประมาณค่ามีน้อย ความแม่นยำของการประมาณค่าก็จะมีมาก ช่วงของค่าประมาณจะแคบ แสดงว่าค่าสารสนเทศมีความสัมพันธ์กับความคลาดเคลื่อนมาตรฐานของการประมาณค่า โดยมีความสัมพันธ์ในทิศทางกลับกัน คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่า = $1/\sqrt{\text{สารสนเทศ}}$

ในทฤษฎีการตอบสนองต่อข้อสอบ จะใช้ผลการตอบแบบสอบประมาณค่าความสามารถของบุคคล การจะประเมินคุณภาพของแบบสอบอาจดูได้จากความถูกต้องแม่นยำในการประมาณค่าความสามารถโดยใช้คะแนนจากแบบสอบ ค่าสารสนเทศจากแบบสอบจะเป็นดัชนีชี้ถึงความถูกต้องแม่นยำของการประมาณค่า (Birnbaum 1968: 418) โดยค่าสารสนเทศคือปริมาณที่เป็นส่วนกลับของกำลังสองของความกว้างของช่วงความเชื่อมั่นของค่าความสามารถด้วยคะแนนจากแบบ

สอบ (Lord 1980: 65) ถ้าให้ y เป็นคะแนนที่ได้จากการตอบแบบสอบ ค่าสารสนเทศสำหรับคะแนน y แสดงได้ด้วยสมการ

$$I\ (\theta\ ,\ y) = [d\mu_{y|\theta}]^2\ /\ Var(y|\theta) \qquad \text{(Lord 1980: 67)}$$

จากสูตรจะเห็นว่าค่าสารสนเทศสำหรับคะแนน y ก็คือ กำลังสองของอัตราส่วนระหว่างความชันของเส้นถดถอยของ y บน θ กับความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถ θ นั่นเอง ดังนั้นค่าสารสนเทศของแบบสอบฉบับหนึ่งมีได้หลายค่าแตกต่างกันไปตามระดับของ θ และที่ระดับความสามารถค่าใดค่าหนึ่ง ค่าสารสนเทศจะมากหรือน้อยขึ้นอยู่กับ

(ก) ค่าความคลาดเคลื่อนมาตรฐานของการวัด y ณ ระดับความสามารถนั้น ซึ่งถ้ามีความคลาดเคลื่อนน้อยค่าสารสนเทศจาก y เกี่ยวกับ θ ก็จะมาก

(ข) ความชันของเส้นถดถอยของ y บน θ ซึ่งถ้ามีความชันมากค่าสารสนเทศจาก y เกี่ยวกับ θ ก็จะมาก (Lord 1980: 68)

นอกจากนี้ค่าสารสนเทศของแบบสอบจะเพิ่มมากขึ้นเมื่อแบบสอบมีข้อสอบจำนวนมากขึ้น (Hambleton, and Swaminathan 1985: 107)

Birnbaum (1968: 453-454) ได้ให้สูตรทั่วไปสำหรับคำนวณค่าสารสนเทศของคะแนน y จากแบบสอบใดๆ ไว้ดังนี้

$$\text{ถ้า} \quad y = \sum_{i=1}^{n} w_i \cdot u_i$$

ค่าสารสนเทศของแบบสอบ คือ

$$I(\theta\ ,\ y) = [\ \sum_{i=1}^{n} w_i\ P_i'(\theta)\ ]^2\ /\ \sum_{i=1}^{n} w_i^2\ P_i(\theta)Q_i(\theta)$$

$$\text{เมื่อ} \qquad P_i'(\theta) = d\ P_i(\theta)\ /\ d\theta$$

ส่วนค่าสารสนเทศของข้อสอบก็จะหาได้จากการแทนค่า y ด้วย u_i นั่นก็คือ w_i มีค่าเท่ากับ 1 และ n มีค่าเท่ากับ 1 สูตรหาค่าสารสนเทศของข้อสอบจึงเป็น

$$I(\Theta, u_i) = [P_i'(\Theta)]^2 / P_i(\Theta)Q_i(\Theta)$$

ในกรณีที่การให้คะแนนโดยการกำหนดน้ำหนักอย่างเหมาะสม (optimal weight) โดย

$$w_i = P_i'(\Theta) / P_i(\Theta)Q_i(\Theta)$$

ค่าสารสนเทศของแบบสอบจะเป็น

$$I(\Theta) = \sum_{i=1}^{n} [\{P_i'(\Theta)\}^2 / P_i(\Theta)Q_i(\Theta)]$$

ซึ่งก็คือผลรวมของค่าสารสนเทศของข้อสอบแต่ละข้อในแบบสอบนั่นเอง ค่า $I(\Theta)$ จะเป็นค่าสารสนเทศที่มากที่สุดที่จะได้จากแบบสอบฉบับใดฉบับหนึ่ง [นั่นคือ $I(\Theta, y) < I(\Theta)$ เสมอ]

จากฟังก์ชันลักษณะของข้อสอบ (item characteristic function) [สมการที่ (4) - (8)] จะเห็นว่า $P_i(\Theta)$ ขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบข้อที่ i คือ a_i, b_i และ c_i ด้วย ดังนั้น ค่าพารามิเตอร์ของข้อสอบจึงเป็นตัวกำหนดค่าสารสนเทศของข้อสอบและของแบบสอบ จึงอาจกล่าวได้ว่า ค่าสารสนเทศของข้อสอบหรือแบบสอบเป็นดัชนีผสม (composite index) ที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลายลักษณะ รวมเป็นดัชนีเพียงตัวเดียวเพื่อชี้ถึงคุณภาพของข้อสอบหรือแบบสอบได้ และด้วยคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ ค่าสารสนเทศจึงเหมาะที่จะใช้เป็นดัชนีบอกคุณภาพของข้อสอบหรือแบบสอบได้ดีกว่าค่าสถิติหรือค่าดัชนีอื่น ๆ ตามแนวคิดของทฤษฎีการวัดแบบคลาสสิค

ประสิทธิภาพของแบบสอบและการเปรียบเทียบ

จุดมุ่งหมายของการสอบตามทฤษฎีการตอบสนองต่อข้อสอบก็คือ ต้องการใช้ผลจากการสอบประมาณค่าความสามารถที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้โดยตรง ดังนั้น การประเมินคุณภาพของแบบสอบจึงต้องพิจารณากันที่ความถูกต้องแม่นยำของการใช้คะแนนสอบประมาณค่าระดับความสามารถ ดัชนีที่บอกถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถก็คือ ค่าสารสนเทศของแบบสอบ เนื่องจากวิธีการให้คะแนนใด ๆ ที่ไม่ใช่การกำหนดน้ำหนักที่เหมาะสม

(optimal scoring weight) แล้ว ค่าสารสนเทศที่ได้จะมีค่าน้อยกว่าค่าสารสนเทศที่ได้จากวิธีการให้คะแนนโดยการกำหนดน้ำหนักคะแนนที่เหมาะสมเสมอในทุกระดับความสามารถ Birnbaum (1968: 471-472) ได้เสนอแนวคิดในการวัดประสิทธิภาพของแบบสอบโดยถือว่าแบบสอบที่มีประสิทธิภาพมากคือแบบสอบที่คะแนนจากการสอบให้ค่าสารสนเทศสูงใกล้เคียงกับสารสนเทศที่ได้จากการให้คะแนนแบบกำหนดน้ำหนักที่เหมาะสม ดัชนีบอกประสิทธิภาพของแบบสอบก็คือ อัตราส่วนระหว่างค่าสารสนเทศที่ได้จากการให้คะแนนแบบ x กับค่าสารสนเทศที่ได้จากการให้คะแนนแบบกำหนดน้ำหนักที่เหมาะสม นั่นคือ

$$\mathrm{Eff}\ (\theta\ ,\ x) \quad = \quad I(\theta\ ,\ x)/I(\theta)$$

ในกรณีที่ต้องการเปรียบเทียบประสิทธิภาพของแบบสอบ 2 ฉบับที่ใช้วัดลักษณะหรือความสามารถอันเดียวกันก็สามารถหาค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์ (relative efficiency) ของแบบสอบทั้งสองได้ โดยหาอัตราส่วนระหว่างค่าสารสนเทศที่ได้จากแบบสอบทั้ง 2 ฉบับ นั่นคือ

$$\mathrm{R.E.}\ (\theta\ ,\ x_1\ ,\ x_2) \quad = \quad I_1(\theta\ ,\ x_1)/I_2(\theta\ ,\ x_2)$$

ค่าดัชนีประสิทธิภาพเชิงสัมพัทธ์นี้ไม่ขึ้นอยู่กับมาตราที่ใช้วัดความสามารถ นั่นคือ ค่าดัชนีจะไม่แปรเปลี่ยนไปตามการแปลงมาตราการวัดความสามารถ (Lord 1980: 89)

การวิจัยเกี่ยวกับการเปรียบเทียบประสิทธิภาพของแบบสอบ

การนำแนวคิดเกี่ยวกับประสิทธิภาพเชิงสัมพัทธ์ไปใช้เปรียบเทียบงประสิทธิภาพของแบบแบบสอบยังไม่เป็นที่แพร่หลายมากนัก จึงไม่ค่อยมีรายงานผลการวิจัยเกี่ยวกับเรื่องนี้มากนัก นอกจากผลงานของ Lord (1974: 247-254) ได้ใช้แนวคิดนี้ไปเปรียบเทียบประสิทธิภาพของแบบสอบการอ่านเมโทรโพลิแทน (Metropolitan Reading Tests) ระดับกลาง (intermediate level) ฟอร์มเอฟ เฉพาะแบบสอบย่อยการวิเคราะห์คำ (word analysis subtest) [MAT] กับแบบสอบการอ่านอื่น ๆ อีก 6 ฉบับ คือ

1. Sequential Test of Educational Progress
   Series II, Level 4, Form A, Reading subtest. [STEP]

2. California Achievement Tests, Level 4, Form A, Reading Vocabulary. [CAT]
3. Iowa Test of Basic Skills, Level 12, Form 5 Vocabulary. [ITBS]
4. Stanford Reading Tests, Intermediate II, Form W Word Meaning. [SRT]
5. Comprehensive Tests of Basic Skills, Level 3, Form Q Reading Vocabulary. [CTBS]
6. SRA Achievement Series, Greén Edition, Form E, Vocabulary. [SRA]

ผลการวิจัยพบว่า

STEP มีประสิทธิภาพสูงกว่า MAT ในกลุ่มนักเรียนที่ความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซนไตล์ที่ 20) และมีประสิทธิภาพต่ำกว่าในระดับความสามารถอื่น ๆ

CAT มีประสิทธิภาพต่ำกว่า MAT เกือบทุกระดับความสามารถ ยกเว้นที่ระดับความสามารถสูงมาก (สูงกว่าตำแหน่งเปอร์เซนไตล์ที่ 95) CAT มีประสิทธิภาพสูงกว่า MAT

ITBS มีประสิทธิภาพต่ำกว่า MAT เกือบทุกระดับความสามารถ ยกเว้นที่ระดับความสามารถสูงมาก (สูงกว่าตำแหน่งเปอร์เซนไตล์ที่ 90) ITBS มีประสิทธิภาพสูงกว่า MAT

SRT มีประสิทธิภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซนไตล์ที่ 50) และมีประสิทธิภาพสูงกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถสูง (สูงกว่าตำแหน่งเปอร์เซนไตล์ที่ 50)

CTBS มีประสิทธิภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซนไตล์ที่ 50) มีประสิทธิภาพใกล้เคียงกันในกลุ่มนักเรียนที่มีความสามารถปานกลางและค่อนข้างสูง และมีประสิทธิภาพสูงกว่าในกลุ่มที่มีความสามารถสูง (สูงกว่าตำแหน่งเปอร์เซนไตล์ที่ 80)

SRA มีประสิทธิภาพต่ำกว่า MAT ในกลุ่มนักเรียนที่มีความสามารถต่ำ (ต่ำกว่าตำแหน่งเปอร์เซนไตล์ที่ 30) และมีประสิทธิภาพสูงกว่าในระดับความสามารถอื่น ๆ

นอกจากเรื่องนี้แล้ว Lord ยังได้ใช้แนวคิดเกี่ยวกับการเปรียบเทียบประสิทธิภาพของแบบสอบด้วยค่าสารสนเทศจากแบบสอบนี้ ไปศึกษาเปรียบเทียบคุณภาพของแบบสอบที่ประกอบด้วย

ข้อสอบเลือกตอบที่มีจำนวนตัวเลือกต่างกัน พบว่า การลดจำนวนตัวเลือกต่อข้อลงในขณะที่เพิ่มจำนวนข้อสอบให้มากขึ้นมีผลทำให้ประสิทธิภาพของแบบสอบเพิ่มขึ้นในกลุ่มผู้ตอบที่มีความสามารถในระดับสูงและทำให้ประสิทธิภาพของแบบสอบลดลงในกลุ่มผู้ตอบที่มีความสามารถในระดับต่ำ และแบบสอบที่ประกอบด้วยข้อสอบที่มีจำนวนตัวเลือก 3 ตัวมีประสิทธิภาพสูงที่สุด (Lord 1976: 36-39)

ประสิทธิภาพของการตัดสินใจของแบบสอบวัดการเรียนเพื่อรู้จริง

ในระบบการเรียนเพื่อรู้จริง (mastery learning) นั้น มีความจำเป็นที่ผู้สอนจะต้องตัดสินใจจำแนกให้ได้ว่าผู้เรียนไหนเป็นผู้รู้จริงแล้ว (master) และผู้เรียนคนไหนยังเป็นผู้ที่ไม่รู้จริง (nonmaster) เครื่องมือที่ผู้สอนมักจะนำมาใช้เพื่อช่วยตัดสินใจก็คือ แบบสอบวัดการเรียนเพื่อรู้จริง (mastery test) ซึ่งเป็นแบบสอบที่นำมาใช้ในจุดมุ่งหมายเพื่อระบุว่าผู้เข้าสอบมีความความสามารถถึงระดับที่กำหนดไว้หรือไม่ (Lord 1980: 162) ประสิทธิภาพของแบบสอบที่ใช้เพื่อการนี้จึงอยู่ที่ว่าคะแนนจากแบบสอบนั้นสามารถใช้ช่วยตัดสินใจได้ถูกต้องเพียงใดหรือการใช้คะแนนจากแบบสอบช่วยตัดสินใจจะช่วยลดความคลาดเคลื่อนจากกการตัดสินใจโดยไม่ใช้คะแนนจากแบบสอบได้มากน้อยเพียงใด Huynh (1980: 385-406) ได้ให้แนวคิดเกี่ยวกับการกำหนดคะแนนจุดตัดและการวัดประสิทธิภาพของการตัดสินใจของแบบสอบ โดยอาศัยแบบจำลองโลจิสต์และทฤษฎีการตัดสินใจเชิงสถิติ (Statistical Decision Theory) ดังนี้

ถ้าสมมุติให้ θ เป็นความสามารถที่ต้องการวัด สำหรับมาตรวัดในหน่วยของ θ จะกำหนดให้ผู้ที่มีความสามารถ $\theta \geq \theta_o$ เป็นผู้รู้จริง (master) ส่วนผู้ที่มีความสามารถ $\theta \leq \theta_o$ จะถือว่าเป็นผู้ที่ไม่รู้จริง (nonmaster) เมื่อ θ_o เป็นค่าคงที่ที่หมายถึงคะแนนจุดตัดที่แท้จริง (true mastery score) ในระบบการเรียนเพื่อรู้จริงผู้สอนจะต้องตัดสินผลการเรียนของผู้เรียนโดยมีทางเลือกอยู่ 2 ทาง คือ ตัดสินว่าผู้เรียนเป็นผู้รู้จริง (a_1) กับตัดสินว่าผู้เรียนเป็นผู้ไม่รู้จริง (a_2) เนื่องจาก θ เป็นลักษณะที่เราไม่สามารถสังเกตได้โดยตรง การตัดสินใจของผู้สอนมักจะใช้คะแนนจากแบบสอบเป็นสารสนเทศช่วยในการตัดสินใจ ถ้า x เป็นคะแนนจากแบบสอบ กฎการตัดสินใจ (decision rule) ก็จะเป็น $a_1 = \{ x \mid x \geq c \}$ และ $a_2 = \{ x \mid x < c \}$ เมื่อ c เป็นคะแนนจุดตัดที่เหมาะสมจากแบบสอบ (test passing score) การตัดสินใจโดยใช้คะแนนจากแบบสอบนี้อาจก่อให้เกิดการตัดสินใจผิดพลาด ซึ่งความ

ผิดพลาดที่จะเกิดขึ้นมีอยู่ 2 กรณีด้วยกันคือ กรณีแรก ความจริงแล้วเขาเป็นผู้ไม่รู้จริง ($\theta < \theta_o$) แต่จากคะแนนสอบเราตัดสินให้เขาเป็นผู้รู้จริง ($x \geq c$) ความคลาดเคลื่อนกรณีนี้เรียกว่า ความผิดพลาดเชิงลบ (fasle negative) อีกกรณีหนึ่งก็คือ ความจริงแล้วเขาเป็นผู้รู้จริง ($\theta \geq \theta_o$) แต่จากคะแนนสอบเราตัดสินให้เขาเป็นผู้ไม่รู้จริง ($x < c$) ความคลาดเคลื่อนในกรณีนี้เรียกว่า ความผิดพลาดเชิงบวก (fasle positive) และถ้า $L_1(\theta)$ และ $L_2(\theta)$ เป็นค่าฟังก์ชันการสูญเสีย (loss function) ที่เกิดจากการเลือกทางเลือก a_1 หรือ a_2 ตามลำดับ ความสูญเสียที่คาดหวังหรือค่าฟังก์ชันความเสี่ยง (risk function) จากการตัดสินใจผิดพลาดโดยใช้คะแนนจุดตัด c ณ θ ใดๆ คือ

$$R(c\ ;\ \theta) \;=\; L_1(\theta)\ P(x \geq c|\theta) + L_2(\theta)\ P(x < c|\theta)$$

ถ้าให้ $C_f(\theta)$ และ $C_s(\theta)$ เป็นค่าเสียโอกาส (oppotunity loss) เนื่องจากการตัดสินใจผิดพลาดในเชิงลบ และ ในเชิงบวกตามลำดับ โดย

$$\begin{aligned}
C_f(\theta) &= L_1(\theta) - L_2(\theta) > 0 \quad \text{เมื่อ } \theta < \theta_o \\
C_s(\theta) &= L_2(\theta) - L_1(\theta) > 0 \quad \text{เมื่อ } \theta \geq \theta_o \\
C_f(\theta) &= 0 \quad \text{เมื่อ } \theta \geq \theta_o \\
C_s(\theta) &= 0 \quad \text{เมื่อ } \theta < \theta_o
\end{aligned}$$

ค่าฟังก์ชันความเสี่ยงอาจเขียนได้ใหม่เป็น

$$R(c\ ;\ \theta) = \begin{cases} C_f(\theta)\ P(x \geq c|\theta) & \text{สำหรับ } \theta < \theta_o \\ C_s(\theta)\ P(x < c|\theta) & \text{สำหรับ } \theta \geq \theta_o \end{cases}$$

ถ้าให้ $$L_1(c) = \sup_{\theta < \theta_o} C_f(\theta)\ P(x \geq c|\theta)$$

และ $$L_2(c) = \sup_{\theta > \theta_o} C_s(\theta)\ P(x < c|\theta)$$

ค่ามากที่สุดของ $R(c\ ;\ \theta)$ สำหรับทุกค่าของ θ คือ

$$M(c) = \max \{L_1(c)\ ,\ L_2(c)\}$$

คะแนนจุดตัดจากแบบสอบที่เหมาะสมตามเกณฑ์มินิแมกซ์ (minimax passing score) ก็คือคะแนน c_o ที่ทำให้ $M(c)$ มีค่าน้อยที่สุด และค่า $M(c)$ ที่น้อยที่สุดนี้จะเป็นค่าความเสี่ยงที่น้อยที่สุดในการตัดสินใจด้วยคะแนนจากแบบสอบที่ใช้คะแนนจุดตัดเท่ากับ c_o จะแทนค่าความเสี่ยงที่น้อยที่สุดนี้ด้วย R_o นั่นคือ ถ้าใช้คะแนนจากแบบสอบที่มีคะแนนจุดตัดเท่ากับ c_o ตัดสินผู้เรียนเป็นผู้รู้จริงหรือไม่รู้จริงจะมีความเสี่ยงต่อการตัดสินใจผิดพลาดเท่ากับ R_o นั่นเอง

ถ้าหากว่าคะแนนจากแบบสอบไม่มีความสัมพันธ์กับ θ เลย

ให้ $C_f^* = \sup_{\theta < \theta_o} C_f(\theta)$ และ $C_s^* = \sup_{\theta > \theta_o} C_s(\theta)$

คะแนนจุดตัดตามเกณฑ์มินิแมกซ์จะคือคะแนน c^* ที่ทำให้

$$C_f^*\ P(x \geq c^*) = C_s^*\ P(x < c^*)$$

หรืออีกนัยหนึ่งการตัดสินผลการเรียนจะเป็นไปอย่างสุ่ม (คือไม่ต้องพิจารณาคะแนนจากผลสอบ) โอกาสที่ผู้สอบจะถูกตัดสินให้เป็นผู้รู้จริงจะเท่ากับ $C_s^* / (C_s^* + C_f^*)$ และโอกาสที่เขาจะถูกตัดสินให้เป็นผู้ไม่รู้จริงจะเท่ากับ $C_f^* / (C_s^* + C_f^*)$ และค่าความเสี่ยงต่อการตัดสินใจผิดพลาดก็จะเป็น

$$R^* = C_f^*\ C_s^* \ /\ (C_f^* + C_s^*)$$

เมื่อพิจารณาเปรียบเทียบการตัดสินใจโดยไม่ใช้คะแนนจากแบบสอบกับการตัดสินใจโดยใช้คะแนนจากแบบสอบแล้ว ความเสี่ยงต่อการตัดสินใจผิดพลาดที่ลดลงจะแสดงถึงประสิทธิภาพของการใช้คะแนนจากแบบสอบช่วยตัดสินใจ ดังนั้นจึงนิยามความหมายของประสิทธิภาพในการตัดสินใจของแบบสอบ (Decision Efficiency , η) ดังนี้

$$\eta = (R^* - R_o)/R^* \quad \text{หรือ} \quad 1 - R_o/R^*$$

ซึ่งจะเป็นฟังก์ชันของค่าเสียโอกาส $C_f(\theta)$, $C_s(\theta)$ และค่าพารามิเตอร์ของข้อสอบ

ในกรณีของฟังก์ชันการสูญเสียเป็นฟังก์ชันที่มีค่าคงที่ (constant loss)

$$C_f(\theta) = \begin{cases} 1 & \text{ถ้า } \theta \leq \theta_o \\ 0 & \text{ถ้า } \theta > \theta_o \end{cases}$$

และ

$$C_s(\theta) = \begin{cases} Q & \text{ถ้า } \theta \geq \theta_o + \epsilon \\ 0 & \text{ถ้า } \theta < \theta_o + \epsilon \end{cases}$$

ถ้า Q คือ อัตราการสูญเสีย (loss ratio) ซึ่งก็คือ อัตราส่วนของความคลาดเคลื่อนเนื่องจากการตัดสินใจผิดพลาดเชิงลบกับความคลาดเคลื่อนเนื่องจากการตัดสินใจผิดพลาดเชิงบวก ซึ่งมีค่าคงที่และบริเวณ $\{\theta \mid \theta_o < \theta < \theta_o + \epsilon\}$ เป็นบริเวณที่ไม่สามารถตัดสินใจได้ กรณีนี้ฟังก์ชันความเสี่ยงจะเขียนได้เป็น

$$R(c\,;\theta) = \begin{cases} P(x \geq c|\theta) & \text{เมื่อ } \theta \leq \theta_o \\ QP(x < c|\theta) & \text{เมื่อ } \theta \geq \theta_o + \epsilon \end{cases}$$

คะแนนจุดตัดตามเกณฑ์มินิแมกซ์ก็คือ คะแนน c_o ที่สอดคล้องกับสมการ

$$P(x > c_o|\theta = \theta_o) = QP(x < c_o|\theta = \theta_o + \epsilon)$$

ในการคำนวณเพื่อประมาณค่าคะแนนจุดตัด c_o และค่าดัชนีประสิทธิภาพของการตัดสินใจนี้ Huynh ได้พัฒนาโปรแกรมคำสั่งสำหรับคำนวณค่าไว้ให้แล้วเรียบร้อย

อย่างไรก็ตามการศึกษาประสิทธิภาพของแบบสอบวัดการเรียนเพื่อรู้จริงตามแนวคิดนี้ยังไม่มีรายงานผลการนำไปประยุกต์ใช้แต่อย่างใดในขณะนี้