การวิเคราะห์ความลำเอียงของข้อสอบ

ฉัตรศิริ ปิยะพิมลสิทธิ์

การวิเคราะห์ความลำเอียงของข้อสอบจะใช้ในการตรวจสอบการทำหน้าที่ของข้อสอบเป็นรายข้อว่าจะทำหน้าที่เหมือนกันในกลุ่มผู้สอบที่ต่างกันหรือไม่ โดยปกติเราจะนิยามกลุ่มของผู้สอบที่แตกต่างกันในเรื่องของเชื้อชาติ เพศ อายุ ประสบการณ์ และอื่น ๆ ในการวิเคราะห์จะใช้เมื่อกลุ่มผู้สอบที่สนใจจะศึกษามีความแตกต่างในระดับของความสามารถ ความรู้ หรือทักษะที่ถูกสอบวัด

มีวิธีการจะตรวจสอบความลำเอียงของข้อสอบได้หลายวิธี โดยจะแบ่งออกเป็น 2 กลุ่มก็คือการตรวจสอบด้วยกระบวนการมาตรฐานเดิม (Classical Approaches) และการตรวจสอบด้วยทฤษฎีการตอบข้อสอบ (Item Response Theory Approaches)

1. การตรวจสอบด้วยกระบวนการมาตรฐานเดิม (Classical Approaches)

1.1 วิธีการแปลงค่าความยาก (Trasformed Item Difficulty Methods)

แองกอฟฟ์ และฟอร์ด (Angoff and Ford) ได้เสนอวิธีการแปลงค่าความยากในการตรวจสอบความลำเอียงของข้อสอบ โดยนิยามข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่เบี่ยงเบนไปจากเส้นแกนหลัก โดยการวิเคราะห์นั้นจะใช้ค่าความยากของข้อสอบแต่ละข้อ จากกลุ่มผู้สอบทั้ง 2 กลุ่ม แล้วนำมาแปลงให้เป็นค่าความยากมาตรฐานที่มีค่าเฉลี่ย 13 และส่วนเบี่ยงเบนมาตรฐาน 4 ด้วยสูตร $\Delta = 13 + 4Z$ จากนั้นนำค่าความยากมาตรฐานของทั้ง 2 กลุ่มมาสร้างเป็นแผนภาพกระจัดกระจาย ถ้าทั้งสองกลุ่มมีความสามารถเท่าเทียมกันแล้วจุดต่าง ๆ จะต้องมีลักษณะเป็นวงรีล้อมรอบเส้นแกนหลัก 45 องศาจากจุด 0 ถ้ากลุ่มผู้สอบมีความสามารถแตกต่างกันจุดต่าง ๆ จะกระจัดกระจายออกไปจากเส้นแกนหลัก 45 องศาจากจุด 0 ดูภาพประกอบ 1

ข้อสอบที่มีความลำเอียงคือข้อสอบที่มีคู่ลำดับอยู่ห่างจากเส้นแกนหลักมาก ๆ ซึ่งระยะห่างระหว่างจุดคู่ลำดับกับเส้นแกนหลักคำนวณได้ด้วยสูตร

$$d_g = \frac{(a\Delta_{g1} - \Delta_{g2} + b)}{\sqrt{a^2 + 1}}$$

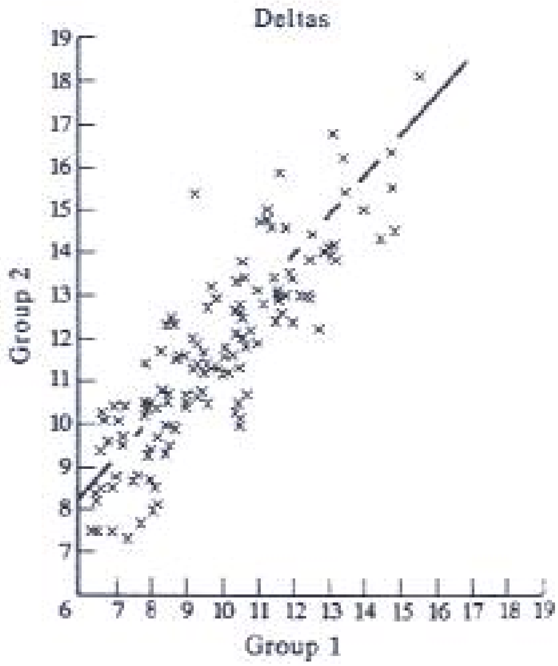
เมื่อ

$$a = \frac{(s_2^2 - s_1^2) + \sqrt{(s_2^2 - s_1^2) + 4r_{12}^2 s_1^2 s_2^2}}{2r_{12}s_1s_2}$$

$$b = \overline{X}_2 - a\overline{X}_1$$

d_g = ระยะห่างระหว่างข้อสอบข้อ g ไปยังเส้นแกนหลัก
Δ_{g1} = ค่าความยากมาตรฐานของข้อสอบข้อ g ในกลุ่ม 1
Δ_{g2} = ค่าความยากมาตรฐานของข้อสอบข้อ g ในกลุ่ม 2
s_1 = ส่วนเบี่ยงเบนมาตรฐานของค่าความยากมาตรฐานกลุ่ม 1
s_2 = ส่วนเบี่ยงเบนมาตรฐานของค่าความยากมาตรฐานกลุ่ม 2
$\bar{X}_1$ = ค่าเฉลี่ยของค่าความยากมาตรฐานในกลุ่ม 1
$\bar{X}_2$ = ค่าเฉลี่ยของค่าความยากมาตรฐานในกลุ่ม 2
r_{12} = สหสัมพันธ์ระหว่างค่าความยากมาตรฐานของกลุ่ม 1 และ 2

ภาพประกอบ 1 แสดงแผนภาพการลงจุดคู่ลำดับค่าความยากมาตรฐานของทั้งสองกลุ่ม



ค่า d_g ของข้อใดมีค่ามากแสดงว่ามีความลำเอียงมาก แต่เกณฑ์ของค่า d_g นี้ยังไม่พบว่ามีเอกสารหรือตำราเล่มใดระบุไว้ว่าค่าเท่าใดถึงจะถือว่าข้อสอบนั้นมีลำเอียง แต่ในงานวิจัยที่ศึกษาเกี่ยวกับความลำเอียงด้วยวิธีการแปลงค่าความยาก จะใช้เกณฑ์ 0.75 หรือก็คือ $d_g < -0.75$ และ $d_g > +0.75$ จึงจะถือว่าข้อสอบข้อนั้นลำเอียง

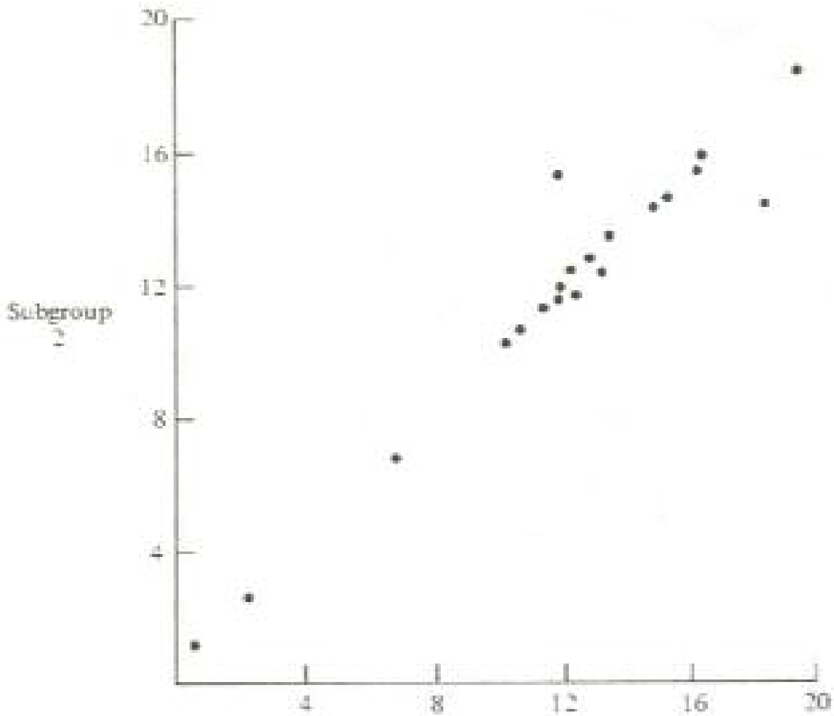
ตัวอย่างคำนวณ

วิธีการหาความลำเอียงด้วยการแปลงค่าความยาก จำเป็นจะต้องใช้ค่าความยากของข้อสอบแปลงเป็นความยากมาตรฐานด้วยสูตร $\Delta_g = 4z_g + 13$ ค่า z_g คือคะแนนมาตรฐาน z ของความยาก (p_g) ที่มีการแจกแจงปกติ และ p_g คือสัดส่วนของการตอบถูกหรือก็คือความยาก

ของข้อสอบ โดยค่า Δ_g จะคำนวณในทุกข้อในทุกกลุ่มและพล็อตกราฟ ตาราง 1 จะแสดงข้อมูลต่าง ๆ ของข้อสอบ 20 ข้อกับกลุ่มตัวอย่าง 2 กลุ่ม สถิติ p_{g1}, z_{g1} และ Δ_{g1} เป็นค่าสถิติของกลุ่มตัวอย่างกลุ่มแรก และ p_{g2}, z_{g2} และ Δ_{g2} เป็นค่าสถิติของกลุ่มสอง ภาพประกอบ 1 แสดงแผนภาพกระจัดกระจายของค่า Δ_{g1} และ Δ_{g2} สหสัมพันธ์ระหว่างค่าความยากมาตรฐานคือ 0.96 ซึ่งเราจะใช้ค่านี้ในการคำนวณ d_g คือระยะห่างของข้อสอบข้อที่ g จากเส้นแกนหลัก จากตัวอย่างนี้ s_1^2 = 20.068, s_2^2 = 17.926, r_{12} = 0.961 และค่า a คำนวณได้ 0.936 จากนั้นคำนวณค่า b ซึ่งมี $\overline{X}_1$ = 12.063, $\overline{X}_2$ = 11.851 และ b จะเท่ากับ 0.558 และแทนค่าในสูตรคำนวณหาระยะห่างจากแกนหลัก โดยข้อที่ 1 คำนวณได้

$$d_1 = \frac{(0.936)(12.058) - (12.466) + (0.558)}{\sqrt{(0.936)^2 + 1}}$$

$$= -0.452$$



ภาพประกอบ 1 แผนภาพกระจัดกระจายของค่าความยากมาตรฐานสองกลุ่ม

ตาราง 1 เปรียบเทียบดัชนีความยากของกลุ่มตัวอย่างสองกลุ่ม

ข้อที่	p_{g1}	z_{g1}	Δ_{g1}	p_{g2}	z_{g2}	Δ_{g2}	d_g
1	0.407	−0.235	12.058	0.447	−0.133	12.466	−0.45
2	0.201	−0.837	9.648	0.226	−0.752	9.988	−0.29
3	0.788	0.802	16.209	0.758	0.703	15.813	−0.06
4	0.430	−0.176	12.293	0.365	−0.344	11.620	0.33
5	0.941	1.566	19.265	0.918	1.394	18.578	0.01
6	0.400	−0.251	11.993	0.404	−0.241	12.036	−0.18
7	0.062	−1.538	6.845	0.065	−1.514	6.942	0.02
8	0.002	−2.828	1.688	0.001	−2.931	1.273	0.63
9	0.390	−0.277	11.888	0.369	−0.332	11.669	0.01
10	0.003	−2.671	2.313	0.005	−2.567	2.730	−0.01
11	0.338	−0.417	11.331	0.351	−0.382	11.470	−0.22
12	0.674	0.452	14.807	0.633	0.341	14.365	0.04
13	0.235	−0.722	10.111	0.183	−0.902	9.391	0.46
14	0.781	0.777	16.108	0.730	0.613	15.453	0.13
15	0.516	0.041	13.166	0.443	−0.143	12.427	0.33
16	0.704	0.538	15.152	0.663	0.423	14.692	0.04
17	0.543	0.108	13.433	0.543	0.109	13.436	−0.22
18	0.488	−0.030	12.877	0.486	−0.034	12.860	−0.18
19	0.387	−0.287	11.850	0.722	0.590	15.360	−2.71
20	0.804	1.305	18.221	0.641	0.362	14.449	2.31

ในตาราง 1 จะสังเกตว่ามีข้อสอบอยู่ 2 ข้อที่มีค่า d_g สูงมากคือข้อที่ 19 และข้อที่ 20 มีค่า d_{19} = -2.71 และ d_{20} = 2.31 นั่นคือข้อสอบทั้ง 2 ข้อนี้มีค่าความความยากมาตรฐานอยู่ห่างจากเส้นแกนหลักมาก นั่นหมายความข้อสอบข้อที่ 19 และ 20 มีความลำเอียง

1.2 วิธีใช้ค่าอำนาจจำแนก (Item Discrimination Procedures)

วิธีนี้เหมือนกับวิธีการแปลงค่าความยาก โดยการคำนวณหาค่าอำนาจจำแนกของข้อสอบของทั้ง 2 กลุ่ม แล้วนำมาสร้างเป็นแผนภาพกระจัดกระจาย แล้วตรวจสอบดูว่าคู่ลำดับของข้อสอบข้อใดที่กระจายออกจากเส้นแกนหลัก 45 องศามากถือว่าข้อสอบข้อนั้นมีความลำเอียง

ส่วน กรีนและดราเปอร์ (Green and Draper) เสนอให้ใช้สหสัมพันธ์พอยท์ไบซีเรียล (Point-biserial correlation) ในการคำนวณหาค่าอำนาจจำแนก แล้วพิจารณาดูว่าข้อสอบข้อใดที่จำแนกได้ดีในกลุ่มหนึ่ง แต่จำแนกได้ไม่ดีในอีกกลุ่มหนึ่ง จะถือว่าข้อสอบข้อนั้นมีความลำเอียง

1.3 วิธีใช้ตารางความสอดคล้อง (Contingency Table Approaches)

วิธีการใช้ตารางความสอดคล้องได้เสนอแนะครั้งแรกโดย ชูนีแมน (Scheuneman) บนพื้นฐานที่ว่า บุคคลที่มีความสามารถเท่ากัน ย่อมมีความน่าจะเป็นในการตอบถูกเท่ากัน ดังนั้นจึงเกิดการสร้างตารางความสอดคล้องระหว่างกลุ่มและความสามารถ ซึ่งจะนิยามเป็นพิสัยของคะแนนในช่วงต่าง ๆ เทคนิคที่ใช้ตารางความสอดคล้องนี้มีหลากหลายดังจะได้อธิบายต่อไปนี้

1.3.1 วิธีไคสแควร์ (Chi-square methods)

ชูนีแมน และคามิลลิ (Scheuneman and Camilli) เทคนิคนี้ได้นิยามข้อสอบที่มีความยุติธรรมไว้ว่า ในแต่ละกลุ่มของผู้สอบที่อยู่ในช่วงคะแนนเดียวกัน สัดส่วนในการตอบถูกของผู้สอบแต่ละกลุ่มจะต้องเท่ากัน

เทคนิคไคสแควร์จะแบ่งคะแนนที่สังเกตได้ออกเป็นช่วง ในแต่ละช่วงจะแสดงสัดส่วนการตอบถูกของผู้สอบในแต่ละกลุ่ม ถ้าสัดส่วนในแต่ละช่วงของแต่ละกลุ่มมีความแตกต่างกันมาก เป็นหลักฐานบ่งชี้ถึงความลำเอียงของข้อสอบ ตาราง 2 แสดงตัวอย่างข้อมูลที่คะแนนแบ่งออกเป็น 4 ช่วง ใช้สัญลักษณ์ N_{1j} และ N_{2j} แทนจำนวนของผู้สอบในกลุ่มแรกและกลุ่มที่สองที่ทำคะแนนได้ในในช่วงที่ j ดังนั้นผู้สอบ 25 คนจากกลุ่มที่ 1 จะได้คะแนนอยู่ในช่วงที่ 1 และ 315 คนจากกลุ่มที่ 2 จะได้คะแนนในช่วงที่ 1 สัญลักษณ์ O_{1j} และ O_{2j} แทนจำนวนของผู้สอบในกลุ่มที่ 1 และกลุ่มที่ 2 ที่ทำคะแนนได้ในช่วงที่ j และตอบข้อสอบข้อนั้นได้ถูก ส่วน P_{1j} จะคำนวณด้วยสูตร

$$P_{1j} = \frac{O_{1j}}{N_{1j}}$$

หรือก็คือสัดส่วนของผู้สอบในกลุ่มที่ 1 และช่วงคะแนนที่ j ที่ตอบข้อสอบข้อนั้นถูก สำหรับกลุ่ม 1 และช่วงคะแนนที่ 3 สัดส่วนในการตอบข้อสอบถูกคือ P_{13} = 23/48 = 0.479 ส่วน P_{2j} ก็คำนวณได้ด้วยสูตรเดียวกัน

และ $P_{.j}$ คำนวณได้จาก

$$P_{.j} = \frac{O_{1j} + O_{2j}}{N_{1j} + N_{2j}}$$

หรือก็คือสัดส่วนของผู้สอบทั้งหมดที่ทำคะแนนได้ในช่วงที่ j และตอบข้อสอบข้อนั้นถูก สำหรับผู้สอบในช่วงคะแนนที่ 4 สัดส่วนของการตอบข้อสอบข้อนั้นถูกคือ

$$P_{.4} = \frac{14 + 33}{65 + 92}$$

$$= \quad 0.299$$

สถิติของคามิลลิ (Camilli's statistic) คำนวณด้วยสูตร

$$\chi_c^2 = \sum_{j=1}^{J} \frac{N_{1j}N_{2j}(P_{1j}-P_{2j})^2}{(N_{1j}+N_{2j})P_{.j}(1-P_{.j})}$$

$$= \Sigma\chi_j^2$$

สถิติ χ_c^2 สามารถทดสอบนัยสำคัญได้ โดยมีองศาแห่งความเป็นอิสระ df = J เมื่อ J ก็คือจำนวนระดับคะแนนที่ถูกแบ่ง ขนาดของไคสแควร์ที่คำนวณได้จะใช้พิจารณาถึงความลำเอียงของข้อสอบ ถ้าไคสแควร์มีค่ามากหรือมีนัยสำคัญทางสถิติแสดงว่าข้อสอบนั้นลำเอียง

ตาราง 2 ข้อมูลที่ใช้ในการคำนวณ χ_c^2 และ χ_s^2

ชั้นที่	ระดับคะแนน	N_{1j}	O_{1j}	P_{1j}	N_{2j}	O_{2j}	P_{2j}	$P_{.j}$
1	13 – 14	25	22	0.880	315	300	0.952	0.947
2	12	24	18	0.750	110	99	0.900	0.873
3	10 – 11	48	23	0.479	118	93	0.788	0.698
4	1 – 9	65	14	0.215	92	33	0.358	0.299

ตาราง 3 แสดงการคำนวณสถิติของคามิลลิ (Camilli's χ_c^2)

ชั้นที่	การคำนวณ χ_j^2	ผลการคำนวณ
1	$\frac{25(315)(0.880-0.952)^2}{(25+315)0.947(1-0.947)}$	2.392
2	$\frac{24(110)(0.750-0.900)^2}{(24+110)0.873(1-0.873)}$	3.998
3	$\frac{48(118)(0.479-0.788)^2}{(48+118)0.698(1-0.698)}$	16.455
4	$\frac{65(92)(0.215-0.358)^2}{(65+92)0.299(1-0.299)}$	3.716
χ_c^2 = 2.392 + 3.998 + 16.445 + 3.716 = 26.56		

สถิติของชูนีแมน (Scheuneman's statistic) คำนวณด้วยสูตร

$$\chi_s^2 = \sum_{j=1}^{J} \frac{(O_{1j}-P_{.j}N_{1j})^2}{P_{.j}N_{1j}} + \sum_{j=1}^{J} \frac{(O_{2j}-P_{.j}N_{2j})^2}{P_{.j}N_{2j}}$$

ตาราง 4 แสดงขั้นตอนการคำนวณ χ_s^2 ชูนีแมนเสนอว่า χ_s^2 คือการแจกแจงไคสแควร์ที่มีองศาแห่งความเป็นอิสระ df = J – 1

ตาราง 4 แสดงการคำนวณสถิติของชูนีแมน (Scheuneman's χ_s^2)

ชั้นที่	กลุ่ม 1	ผลการคำนวณ	กลุ่ม 2	ผลการคำนวณ
1	$\frac{[22-(0.947)(25)]^2}{(0.947)(25)}$	0.118	$\frac{[300-(0.947)(315)]^2}{(0.947)(315)}$	0.009
2	$\frac{[18-(0.873)(24)]^2}{(0.879)(24)}$	0.415	$\frac{[99-(0.873)(110)]^2}{(0.873)(110)}$	0.091
3	$\frac{[23-(0.698)(48)]^2}{(0.698)(48)}$	3.293	$\frac{[93-(0.698)(118)]^2}{(0.698)(118)}$	1.373
4	$\frac{[14-(0.299)(65)]^2}{(0.299)(65)}$	1.519	$\frac{[33-(0.299)(92)]^2}{(0.299)(92)}$	1.096
χ_s^2 = 0.118 + 0.415 + 3.293 + 1.519 + 0.009 + 0.091 + 1.373 + 1.096 = 7.89				

ประเด็นปัญหาของวิธีการไคสแควร์ก็คือการเลือกคะแนนจุดตัดในการแบ่งคะแนนออกเป็นช่วง ๆ การตัดสินใจเกี่ยวกับคะแนนจุดตัดมีความสำคัญมากต่อขนาดของ χ_c^2 และ χ_s^2 ชูนีแมน (Scheuneman) ได้เสนอแนะว่า การคำนวณด้วยสูตร χ_s^2 จะต้องแบ่งคะแนนออกเป็นช่วงโดยที่ $N_{ij}P_{.j}$ มีค่าอย่างน้อย 5 ในทุก i และทุก j ส่วน อิรอนสัน (Ironson) ได้เสนอแนะว่า การคำนวณด้วยสูตร χ_c^2 จะต้องแบ่งคะแนนออกเป็นช่วงโดยที่ $N_{ij}P_{.j}$ และ $N_{ij}(1 - P_{.j})$ ควรมีค่าอย่างน้อย 5 เช่นเดียวกัน

1.3.2 วิธีเมนเทล-เฮนเซล (Mantel-Haenszel procedure)

วิธีนี้พัฒนาขึ้นโดย เมนเทลและเฮนเซล (Mantel and Haenszel) ซึ่งกระบวนการนี้นิยมใช้ในด้านทางการแพทย์ ต่อมาฮอลแลนด์และเทเยอร์ (Holland and Thayer) ได้นำมาประยุกต์ใช้ในการหาความลำเอียงของข้อสอบ มีสูตรว่า

$$\alpha_{MH} = \frac{\sum_j A_j D_j / T_j}{\sum_j B_j C_j / T_j}$$

เมื่อ A_j แทนจำนวนผู้ตอบถูกในกลุ่ม 1 ที่ระดับคะแนน j

B_j แทนจำนวนผู้ตอบผิดในกลุ่ม 1 ที่ระดับคะแนน j

C_j แทนจำนวนผู้ตอบถูกในกลุ่ม 2 ที่ระดับคะแนน j

D_j แทนจำนวนผู้ตอบผิดในกลุ่ม 2 ที่ระดับคะแนน j

T_j แทนจำนวนผู้ตอบข้อสอบทั้งหมดที่ระดับคะแนน j

ถ้าค่า α_{MH} มีค่าเท่ากับ 1.00 แสดงว่าทั้ง 2 กลุ่มมีความเท่าเทียม แต่ถ้า α_{MH} มีค่ามากกว่า 1.00 แสดงว่ากลุ่ม 1 ตอบข้อสอบได้ถูกต้องมากกว่า และถ้า α_{MH} มีค่าน้อยกว่า 1.00 แสดงว่ากลุ่ม 2 ตอบข้อสอบได้ถูกต้องมากกว่า ในการทดสอบสมมติฐานเกี่ยวกับความเป็นอิสระในการตอบข้อสอบและความเป็นสมาชิกของกลุ่มจะใช้ไคสแควร์ที่มีองศาแห่งความเป็นอิสระ df = 1 ในการประมาณค่าสถิติเมนเทล-เฮนเซล คำนวณได้ด้วยสูตร

$$\chi^2_{MH} = \frac{\left(\left|\sum_j A_j - E(A_j)\right| / 0.5\right)^2}{\sum_j S(A_j)^2}$$

เมื่อ $E(A_j) = [(A_j + B_j)(A_j + C_j)] / T_j$

และ $S(A_j)^2 = \dfrac{(A_j + B_j)(C_j + D_j)(A_j + C_j)(B_j + D_j)}{T_j^2 (T_j - 1)}$

วิธีนี้ต้องการกลุ่มตัวอย่างขนาดใหญ่กว่าวิธีอื่น ๆ (ประมาณ 500 คนต่อกลุ่ม) และยังไม่ต้องใช้โปรแกรมคอมพิวเตอร์ในการคำนวณเหมือนกับวิธีของทฤษฎีการตอบข้อสอบ ถ้าไม่แน่ใจว่าข้อสอบมีความลำเอียง วิธีเมนเทล-เฮนเซลก็ไม่เหมาะที่จะใช้เพราะอาจจะมีความผิดพลาดในการค้นหาความลำเอียงของข้อสอบ

1.4 วิธีมาตรฐาน (Standardization Procedure)

วิธีนี้พัฒนาขึ้นมาโดย โดเรนและคูลิค (Dorans and Kulick) บนพื้นฐานของฟังก์ชันการตอบข้อสอบเชิงประจักษ์เมื่อความน่าจะเป็นของการตอบข้อสอบถูกถูกประมาณค่าโดยการสังเกตสัดส่วนการตอบถูกในแต่ละระดับความสามารถ การประมาณค่าความน่าจะเป็นตามเงื่อนไขของความสำเร็จในแต่ละระดับคะแนนจะได้มาจาก “กลุ่มอ้างอิง” (reference group) เป็นกลุ่มที่มีรูปแบบการตอบเพื่อให้ “กลุ่มสนใจ” (focal group) ได้อ้างอิง โดยปกติกลุ่มสนใจจะเป็นกลุ่มที่มีคะแนนต่ำกว่า) กลุ่มอ้างอิงโดยปกติจะมีขนาดใหญ่กว่ากลุ่มสนใจและมีการประมาณค่าพิสัยของคะแนนที่แน่นอน

ดัชนีความลำเอียง (DIF index) จะคำนวณด้วยวิธีมาตรฐานที่ใช้ฟังก์ชันการถ่วงน้ำหนักในทางปฏิบัติ การเลือกน้ำหนักจะขึ้นอยู่กับจำนวนของผู้สอบในกลุ่มสนใจที่ระดับคะแนน (K_j) เพราะน้ำหนักนี้มีความแตกต่างระหว่างสัดส่วนของการตอบถูกสำหรับกลุ่มอ้างอิงและกลุ่มสนใจจะถ่วงน้ำหนักมากที่ระดับคะแนนที่สมาชิกกลุ่มสนใจส่วนมากทำได้ ดัชนีความแตกต่างของข้อสอบจะเป็นความแตกต่างของค่า p มาตรฐาน ดังสูตรว่า

$$\text{Dstd} = \frac{\sum_j K_j (P_{jf} - p_{jr})}{\sum_j K_j}$$

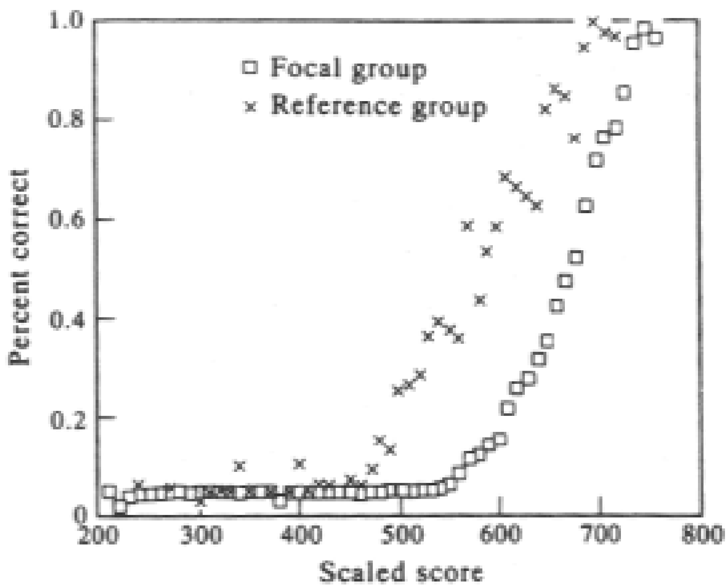
เมื่อ $(K_j/\Sigma K_j)$ แทน องค์ประกอบที่ถูกถ่วงน้ำหนัก

P_{jf} แทน ความน่าจะเป็นของสมาชิกกลุ่มสนใจที่ระดับคะแนนนั้นตอบข้อสอบถูกต้อง

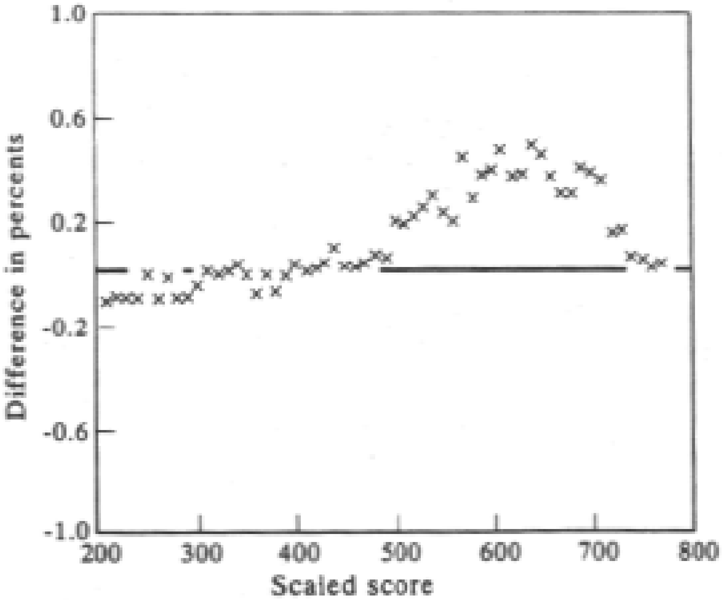
P_{jr} แทน ความน่าจะเป็นของสมาชิกกลุ่มอ้างอิงที่ระดับคะแนนนั้นตอบข้อสอบถูกต้อง

ข้อดีของวิธีนี้คือมีการแสดงแผนภาพของความน่าจะเป็นตามเงื่อนไขของความสำเร็จในการตอบข้อสอบและความแตกต่างระหว่างความน่าจะเป็นของกลุ่มสนใจและกลุ่มอ้างอิงสำหรับข้อสอบแต่ละข้อแสดงดังภาพประกอบ 3 และ 4 ในที่นี้จะแสดงแผนภาพเฉพาะข้อสอบที่มีค่า Dstd สูงเท่านั้น

การคำนวณด้วยวิธีมาตรฐานจะคล้ายคลึงกับวิธีเมนเทล-เฮนเซล แตกต่างกันตรงการใช้ถ่วงน้ำหนัก แต่วิธีมาตรฐานจะประมาณค่าได้ดีกว่า ส่วนข้อด้อยของวิธีทั้งสองก็คือจะต้องใช้กลุ่มตัวอย่างขนาดใหญ่จึงจะประมาณค่าได้แม่นยำ



ภาพประกอบ 3 ความน่าจะเป็นตามเงื่อนไขของการตอบถูกในกลุ่มสนใจและกลุ่มอ้างอิง



ภาพประกอบ 4 ความแตกต่างของความน่าจะเป็นตามเงื่อนไขของการตอบถูก ในกลุ่มสนใจและกลุ่มอ้างอิง

1.5 การวิเคราะห์ตัวลวง (Distractor Analysis)

การวิเคราะห์หาความลำเอียงโดยการใช้ตัวลวงนั้น ชูนีแมน (Scheuneman) ได้เสนอแนะวิธีการวิเคราะห์ตัวลวงไว้ 3 วิธีก็คือ 1) เปรียบเทียบคะแนนเฉลี่ยของแต่ละบุคคลกับตัวลวงแต่ละตัว 2) วีลและโฟร์แมน (Veale and Forman) ได้ใช้ไคสแควร์ในการวิเคราะห์ตัวลวงเท่านั้น 3) ฟรารี่และกิล (Frary and Giles) ได้ใช้วิธีของราสช์ (Rasch model method)

ส่วน กรีนและคนอื่น (Green and other) ได้ใช้วิธีล็อกลิเนียร์ในการตรวจสอบความน่าจะเป็นของปฏิบัติสัมพันธ์ระหว่างกลุ่มกับตัวลวงเมื่อควบคุมความสามารถให้คงที่ นอกจากนี้ยังสามารถประยุกต์ตารางความสอดคล้องมาใช้ในการวิเคราะห์ตัวลวงได้อีก โดยกลุ่มตัวอย่างควรมีขนาดใหญ่เพียงพอ การวิเคราะห์ตัวลวงเพียงหนึ่งตัวหรือมากกว่าหนึ่งตัวน่าจะดีกว่าการวิเคราะห์ตัวเลือกถูกเพียงตัวเดียว

วิธีการวิเคราะห์ตัวลวง

ขั้นแรกในการวิเคราะห์ตัวลวงเพื่อใช้ในการตรวจสอบความลำเอียงข้อสอบจะต้องเตรียมเมตริกของตัวเลือกทั้งหมดของข้อสอบ ดังเช่นตาราง 4 ที่ศึกษากับกลุ่ม 3 กลุ่ม และข้อสอบมี 4 ตัวเลือก ตัวเลือกถูกของข้อสอบข้อนี้คือ C และตัวลวงอีก 3 ตัวคือ A, B และ D สังเกตว่าตาราง 5 ไม่ได้แสดงเฉพาะตัวลวงเท่านั้น ยังแสดงตัวถูกไว้ด้วย และยังแสดงจำนวนของผู้สอบที่เลือกสอบ 2 ตัวเลือกเอาไว้ด้วย ขั้นตอนถัดไปจะแทนที่ข้อมูลในชุดของตาราง 2 x k เพื่อเตรียมทดสอบนัยสำคัญ ตาราง 6 จะแสดงข้อมูลที่นำมาจากตาราง 5

ตาราง 5 เมตริกการเลือกตอบตัวเลือกของข้อสอบข้อหนึ่ง

กลุ่ม	ตัวเลือก A	B	C*	D	Double Marks	Omits	รวม
1	32	32	56	16	12	0	136
2	24	72	40	24	16	0	176
3	24	5	8	6	4	0	47

วีลและโฟร์แมน (Veale and Forman. 1975) ได้เสนอวิธีการวิเคราะห์ตัวลวงด้วยการใช้ไคสแควร์ในการทดสอบ ดังผลการทดสอบในตาราง 6

ตาราง 6 ตารางความสอดคล้องสำหรับทั้ง 3 กลุ่มในการวิเคราะห์ตัวลวงของข้อสอบข้อหนึ่ง

ตัวเลือก A

กลุ่ม	1	2	3
ตัวถูก	49.74 56	36.17 40	18.09 8
0	38.26 32	27.83 24	13.91 24

χ^2_A = 15.68 ($p < 0.01$)

ตัวเลือก D

กลุ่ม	1	2	3
ตัวถูก	49.92 56	44.37 40	9.71 8
0	22.08 16	19.63 24	4.29 6

χ^2_D = 4.80

ตัวเลือก B

กลุ่ม	1	2	3
ตัวถูก	42.97 56	54.69 40	6.35 8
0	45.03 32	57.31 72	6.65 5

χ^2_B = 16.27 ($p < 0.01$)

Double Marks

กลุ่ม	1	2	3
ตัวถูก	52.00 56	42.82 40	9.18 8
0	16.00 12	13.18 16	2.82 4

χ^2_{DM} = 2.74

ระดับนัยสำคัญที่เกิดขึ้นจากการแบ่งส่วนในการทดสอบนั้นย่อมเกิด Type I error มากขึ้น ดังนั้นจึงจำเป็นต้องใช้ระดับนัยสำคัญโดยรวมเพื่อควบคุมการเกิด Type I error ด้วยสูตร

$$\alpha = \frac{\alpha_t}{b}$$

ในการวิเคราะห์ตัวลวงของตัวอย่างนี้ 3 ตัว รวมกับการวิเคราะห์การเลือกตอบตัวเลือกซ้ำ (Double Marks) รวมเป็น 4 ดังนั้นระดับนัยสำคัญโดยรวมที่ 0.05 จะเท่ากับ

$$\alpha = \frac{0.05}{4} = 0.0125$$

ดังนั้นค่าวิกฤติของการทดสอบไคสแควร์ของตัวลวงจะมีค่าระดับนัยสำคัญที่ 0.0125 และ df = 2 จะได้ค่า 9.21 ดังนั้นความลำเอียงของข้อสอบจะพิจารณาจากค่าไคสแควร์ของตัวลวงที่มีค่ามากกว่า 9.21 ซึ่งในตัวอย่างนี้ข้อสอบข้อนี้จะลำเอียงที่ตัวลวง A และ B

แต่สำหรับวิธีการนี้การเลือกใช้ไคสแควร์จะต้องพิจารณาถึงความเหมาะสมด้วย เช่น กลุ่มที่ศึกษาความลำเอียงมีเพียง 2 กลุ่ม df = 1 และความถี่ที่คาดหวังของแต่ละเซลน้อยกว่า 10 ก็ควรจะใช้สูตรปรับแก้ของเยส (Yates's Correction)

1.6 วิธีวิเคราะห์องค์ประกอบ

วิธีนี้จะถือว่าข้อสอบที่มีค่าน้ำหนักองค์ประกอบไม่เท่ากัน เมื่อคำนวณจากกลุ่มผู้สอบที่ต่างกลุ่มกันจะเป็นข้อสอบที่มีความลำเอียง หมายความว่าข้อสอบข้อนั้นไม่ได้วัดความสามารถเดียวกันในผู้สอบที่ต่างกลุ่มกัน

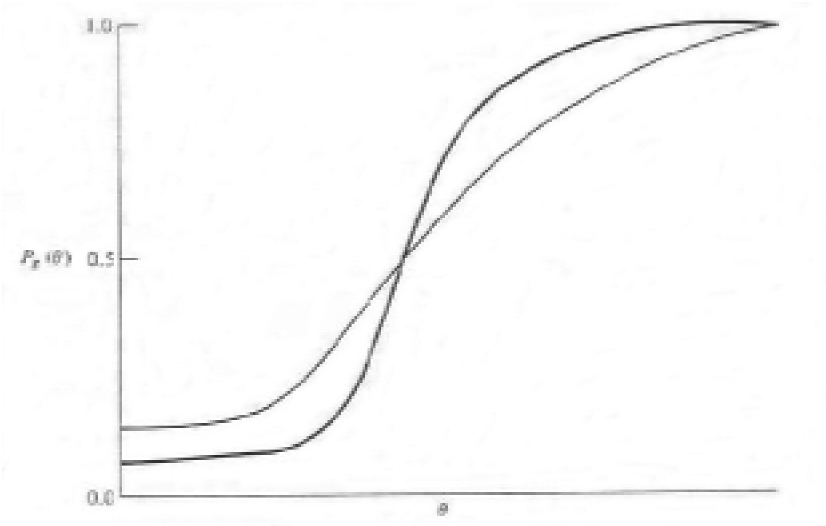
วิธีนี้จะนำข้อสอบมาวิเคราะห์องค์ประกอบโดยวิเคราะห์แยกในแต่ละกลุ่ม แล้วนำค่าน้ำหนักองค์ประกอบที่ได้ในแต่ละข้อมาเปรียบเทียบกัน การเปรียบเทียบจะต้องนำค่าน้ำหนักองค์ประกอบไปแปลงเป็นฟิชเชอร์ซี (Fisher's Z) แล้วนำค่าที่ได้มาทดสอบความแตกต่างด้วยไคสแควร์ หาค่าไคสแควร์มีนัยสำคัญทางสถิติถือว่าข้อสอบข้อนั้นมีความลำเอียง

2. การศึกษาความลำเอียงตามทฤษฎีการตอบข้อสอบ

ในการใช้ทฤษฎีการตอบข้อสอบในการตรวจสอบความลำเอียงของข้อสอบ ชุดของข้อสอบจะถูกตัดสินว่าไม่ลำเอียงถ้าเส้น ICC ของข้อสอบเหมือนกันในกลุ่มตัวอย่างทั้งสองกลุ่ม แล้วผู้สอบจะมีคะแนนคุณลักษณะแฝงเหมือนกัน ข้อสอบจะมีความยากง่ายเท่ากันสำหรับสมาชิกของกลุ่มตัวอย่างทั้งสองกลุ่ม ดังนั้นแหล่งของความแปรปรวนที่มีอิทธิพลต่อกลุ่มตัวอย่างทั้งสองกลุ่มจะเกิดขึ้นเหมือนกัน นอกจากนั้น ICC จะเหมือนกันในทั้งสองกลุ่ม ข้อสอบสอบวัดคุณลักษณะเดียวกันในแต่ละกลุ่ม ดังนั้น ถ้าชุดของข้อสอบไม่ลำเอียงตามนิยามของคุณลักษณะแฝง มักจะไม่ลำเอียงตามนิยามอื่น แต่ข้อสอบที่ลำเอียงในนิยามของคุณลักษณะแฝงอาจจะลำเอียงในนิยามอื่นหรือไม่ก็ได้ ภาพประกอบ 4 จะแสดงข้อสอบสำหรับ ICC ที่ไม่เหมือนกันในทั้งสองกลุ่ม

เพราะว่า ICC สำหรับชุดของข้อสอบที่ไม่ลำเอียงจะมีเปลี่ยนแปลงในทั้งสองกลุ่มเป็นการบ่งชี้ว่าบนพื้นฐานของทฤษฎีการตอบข้อสอบในการวัดโดยใช้ ICC เปรียบเทียบทั้งสองกลุ่มในการคำนวณจะบ่งชี้ถึงความเกี่ยวข้องในสองขั้นตอน ขั้นตอนแรกพารามิเตอร์ข้อสอบจะถูก

ประเมินสำหรับกลุ่มตัวอย่างแต่ละกลุ่มและแสดงบนสเกลที่เหมือนกัน ขั้นที่สองดัชนีของความลำเอียงจะคำนวณแสดงในแต่ละข้อ



ภาพประกอบ 5 แสดงเส้น ICC ที่ไม่เท่ากันในกลุ่มตัวอย่าง 2 กลุ่ม

สเกลของพารามิเตอร์ข้อสอบ

ก่อนคำนวณดัชนีความลำเอียงของข้อสอบ การประมาณค่าพารามิเตอร์ของข้อสอบต้องแสดงบนสเกลเดียวกันสำหรับกลุ่มตัวอย่างแต่ละกลุ่ม วิธีง่าย ๆ วิธีหนึ่งนั่นคือการควบคุมการประมาณค่าของ b_g โดยให้มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 นั่นคือรู้ว่า “มีความเป็นมาตรฐานบน b_g” และแทนที่ด้วยการประมาณค่าพารามิเตอร์บนสเกลที่เหมือนกันในแต่ละกลุ่ม สำหรับกระบวนการของโมเดลโลจิสติก 3 พารามิเตอร์และสำหรับโมเดลนอร์มอนออไจส์ (สำหรับโมเดลพารามิเตอร์เดียวมันจำเป็นที่จะต้องบังคับค่าเฉลี่ยของค่าความยากให้เป็น 0)

อีกทางเลือกหนึ่งจะเกี่ยวข้องกับความเป็นมาตรฐานบนคะแนนคุณลักษณะแฝง (θ) ในทางเลือกนี้ สเกลสำหรับประมาณค่าของคะแนนคุณลักษณะแฝงเป็นชุดที่คำนวณแยกสำหรับแต่ละกลุ่ม ในแต่ละกลุ่มคะแนนคุณลักษณะแฝงจะประมาณค่าโดยมีค่าเฉลี่ย 0 และส่วนเบี่ยงเบนมาตรฐาน 1 สเกลสำหรับคะแนนคุณลักษณะแฝงประมาณค่าแตกต่างกันในทั้งสองกลุ่ม รวมทั้งสเกลสำหรับการประมาณค่าความยากและอำนาจจำแนก อย่างไรก็ตาม มันเป็นไปได้ในการแปลงค่าความยากที่ถูกประมาณค่าและอำนาจจำแนกในกลุ่มที่ 1 ไปเป็นสเกลในกลุ่มที่ 2 โดยใช้สมการ

$$\hat{b}_{1g} = k\hat{b}^*_{1g} + m \qquad (1a)$$

และ

$$\hat{a}_{1g} = \frac{\hat{a}^*_{1g}}{k} \qquad (1b)$$

ในสมการ $\hat{a}_{1g}$ และ $\hat{b}_{1g}$ แสดงการประมาณค่าอำนาจจำแนกและความยากสำหรับกลุ่มตัวอย่างที่ 1 ที่ค่ามาตรฐาน θ และ k และ m เป็นค่าคงที่ในการประมาณค่าที่จะอธิบายต่อมา ค่าของ $\hat{a}_{1g}$ และ $\hat{b}_{1g}$ อยู่บนสเกลที่เหมือนกับ $\hat{a}_{2g}$ และ $\hat{b}_{2g}$ การประมาณค่าจะอ้างอิงกลุ่มตัวอย่างที่ 2 ที่ค่ามาตรฐานθ ปริมาณของ k สามารถประมาณค่าโดยความชันของแกนหลักที่นำค่าของ $\hat{b}_{2g}$ มาพล็อตกับ $\hat{b}_{1g}$ ความชันและจุดตัดของเส้นตรงคำนวณได้ด้วย

$$\hat{k} = \frac{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2) + \sqrt{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2) + 4\hat{\rho}_{12}^2\hat{\sigma}_1^2\hat{\sigma}_2^2}}{2\hat{\rho}_{12}^2\hat{\sigma}_1^2\hat{\sigma}_2^2} \qquad (2a)$$

และ

$$\hat{m} = \overline{b}_2 - \hat{k}\overline{b}_1 \qquad (2b)$$

ตามลำดับ ในสมการ 2a, $\hat{\sigma}_1^2$ แสดงความแปรปรวนของข้อสอบข้อที่ b_{1g}^*, $\hat{\sigma}_2^2$ แสดงความแปรปรวนของ $\hat{b}_{2g}$, และ $\hat{\rho}_{12}$ แสดงความสัมพันธ์ของข้อสอบข้อที่ b_{1g}^* และ $\hat{b}_{2g}$ ในสมการ 2b, b_2 และ b_1 แสดงค่าเฉลี่ยของข้อสอบข้อที่ $\hat{b}_{2g}$ และ b_{1g}^* ตามลำดับ (Ironson 1982) กระบวนการคำนวณนี้จะใช้ได้สำหรับโมเดล 1 และ 2 และ 3 พารามิเตอร์ สำหรับโมเดล 1 พารามิเตอร์ m จะเท่ากับ 1 สังเกตว่ามีพารามิเตอร์การเดา c_g จะไม่มีอิทธิพลต่อการเปลี่ยนแปลงของสเกลและไม่จำเป็นต้องเท่ากับ $\hat{c}'_s$ สำหรับกลุ่มตัวอย่างทั้งสองกลุ่ม

การเปรียบเทียบ ICC

ในการประมาณค่าพารามิเตอร์จะแสดงอยู่บนสเกลเดียวกัน ICC สำหรับทั้ง 2 กลุ่มตัวอย่างจะถูกเปรียบเทียบกัน มีหลายวิธีที่เป็นไปได้สำหรับใช้ในการเปรียบเทียบ เมื่อใช้โมเดลโลจิสติก 3 พารามิเตอร์ Lord (1980) เห็นว่าควรมีการทดสอบสมมติฐานว่า $a_{1g} = a_{2g}$ และ $b_{1g} = b_{2g}$ หรือไม่ สังเกตว่าการทดสอบนี้ใช้ร่วมกับกระบวนการประมาณค่าความเป็นมาตรฐานของ b_g ในกลุ่ม เพราะว่า c_g มีความลำบากที่จะประมาณค่าได้ถูกต้องและไม่รวมอยู่ในการทดสอบ การทดสอบจะทำให้เสร็จในแต่ละข้อและนำไปสู่สถิติไคสแควร์ซึ่งสามารถใช้ในการทดสอบสมติฐานข้อสอบแต่ละข้อ สำหรับสมมติฐานที่ถูกปฏิเสธอาจจะพิจารณาว่าเป็นข้อสอบที่ลำเอียง อีกทางเลือกหนึ่งคือขนาดของสถิติไคสแควร์ที่อาจจะพิจารณาการวัดระดับความลำเอียงของข้อสอบ Lord (1980) ได้พัฒนาการทดสอบด้วยโมเดล 3 พารามิเตอร์ อย่างไรก็ตาม การทดสอบนี้สามารถใช้ได้สำหรับโมเดล 2 พารามิเตอร์ และโมเดล 1 พารามิเตอร์ สมมติฐานศูนย์ $b_{1g} = b_{2g}$ อาจจะทดสอบกับ

$$z = \frac{\hat{b}_{1g} - \hat{b}_{2g}}{\sqrt{\hat{\sigma}_{b_{1g}}^2 + \hat{\sigma}_{b_{2g}}^2}}$$

เมื่อ $\hat{\sigma}^2_{b_{ig}}$ แสดงความคลาดเคลื่อนมาตรฐานของความยากสำหรับข้อสอบข้อที่ g ในกลุ่มที่ i ค่าวิกฤติคือ $\pm z_{\alpha/2}$, เมื่อ $z_{\alpha/2}$ คือ $100(1 - \alpha/2)$ เปอร์เซ็นต์ของการแจกแจงปกติ

Rudner (1977) ได้เสนอให้คำนวณพื้นที่ระหว่าง ICC ของทั้งสองกลุ่ม เป็นการวัดความแตกต่างระหว่าง ICC 2 เส้น พื้นที่นี้สามารถประมาณค่าโดยใช้สูตร

$$A_g = \sum_{\theta=-4.00}^{\theta=4.00} 0.005\left|P_{1g}(\theta) - P_{2g}(\theta)\right| \qquad (3)$$

ในสมการ 3 นี้ $P_{1g}(\theta)$ และ $P_{2g}(\theta)$ อ้างอิงค่าของ ICC ในแต่ละกลุ่ม ค่าของ $P_{ig}(\theta)$ ถูกคำนวณสำหรับแต่ละค่าของ θ จาก –4.00 ถึง 4.00 โดยต่างกัน 0.005 ในแต่ละขั้น สมการ 3 แปรเปลี่ยนไปกลายเป็น

$$A_g(\text{signed}) = \sum_{\theta=-4}^{\theta=4} 0.005\left[P_{1g}(\theta) - P_{2g}(\theta)\right] \qquad (4)$$

ซึ่งได้ผลเป็นเครื่องหมายของการวัดพื้นที่ จำนวนที่แตกต่างระหว่างมีเครื่องหมายและไม่มีเครื่องหมายในการวัดเกิดขึ้นได้เมื่อ ICC สำหรับทั้งสองกลุ่มไขว้กัน แสดงให้เห็นในภาพประกอบ 5 เพราะว่าค่าสัมบูรณ์ของเครื่องหมายในสมการ 3 จะเปลี่ยนค่าลบของ $P_{1g}(\theta) - P_{2g}(\theta)$ ไปเป็นค่าบวก ดังนั้นความแตกต่างทั้งหมดระหว่าง ICC ของทั้งสองกลุ่มตัวอย่างจะให้ค่าเป็นบวกด้วยสมการ 3 อย่างไรก็ตาม สมการ 4 ไม่ได้แปลงค่าลบของ $P_{1g}(\theta) - P_{2g}(\theta)$ เป็นค่าบวก ดังนั้นบวกและลบที่แตกต่างกันจะสมดุลย์กันในบางช่วง และในการวัดพื้นที่ที่ไม่มีเครื่องหมายจะมากกว่าการวัดพื้นที่ที่มีเครื่องหมาย ผลที่ได้คือข้อสอบที่มี ICC เหมือนกันและข้อสอบที่มีเครื่องหมายแตกต่างกันที่สมดุลกันผลก็คือ A_g(signed) = 0

ตัวอย่างการคำนวณโดยใช้ทฤษฎีการตอบข้อสอบ

การแสดงการใช้ทฤษฎีการตอบข้อสอบในการตรวจสอบความลำเอียงของข้อสอบ ด้วยกลุ่มตัวอย่างที่เป็นเพศชาย 150 คนและเพศหญิง 150 คน กับข้อสอบ 10 ข้อที่นำมาใช้ในการวิเคราะห์ ตาราง 7 รายงานการประมาณค่าพารามิเตอร์ความยากง่ายและอำนาจจำแนกของข้อสอบด้วยโมเดลนอร์มอนออไจสำหรับกลุ่มตัวอย่างเพศชายและเพศหญิง ค่าของ $\hat{b}^*_{1g}$ และ $\hat{a}^*_{1g}$ เป็นของกลุ่มตัวอย่างเพศชายและอ้างอิงความเป็นมาตรฐานด้วยค่า θ สำหรับกลุ่มเพศชาย ค่าของ $\hat{b}_{2g}$ และ $\hat{a}_{2g}$ เป็นของกลุ่มตัวอย่างเพศหญิงและอ้างอิงความเป็นมาตรฐานด้วยค่า θ สำหรับกลุ่มเพศหญิง ตั้งแต่การประมาณค่าพารามิเตอร์สำหรับเพศชายและสำหรับเพศหญิงจะถูกแสดงบนสเกลที่แตกต่างกัน ดังนั้นจึงไม่สามารถนำมาเปรียบเทียบกันได้โดยตรง ต้องทำการให้สามารถเปรียบเทียบกันได้ $\hat{b}^*_{1g}$ และ $\hat{a}^*_{1g}$ ต้องถูกแปลงด้วยสมการ 1a และ 1b นั่นคือต้องคำนวณ $\hat{k}$ และ $\hat{m}$ ตาราง 7 รายงาน $\bar{b}_1$, $\hat{\sigma}^2_1$, $\bar{b}_2$, $\hat{\sigma}^2_2$ และ $\hat{\rho}_{12}$ ซึ่งจำเป็นสำหรับคำนวณ $\hat{k}$ และ $\hat{m}$ แทนที่ค่าในสมการ 2a จะได้

$$\hat{k} = \frac{(1.055-1.343)+\sqrt{(1.055-1.343)+4(0.769)^2(1.055)(1.343)}}{2(0.769)(1.027)(1.158)}$$

$$= 0.798$$

ตาราง 7 การประมาณค่าความยากง่ายและอำนาจจำแนกสำหรับกลุ่มตัวอย่างเพศชายและหญิง

ข้อสอบ	ชาย		หญิง	
	$\hat{b}^*_{1g}$	$\hat{a}^*_{1g}$	$\hat{b}_{2g}$	$\hat{a}_{2g}$
1	-3.726	0.527	-1.467	1.198
2	-0.115	1.000	0.222	0.723
3	-1.479	2.528	-1.602	0.989
4	-2.966	0.424	-1.734	0.592
5	-2.614	0.571	-3.202	0.589
6	-1.375	1.730	-1.278	0.909
7	-0.633	0.881	0.139	0.718
8	-1.196	0.928	-0.428	0.963
9	-0.943	1.920	-0.624	0.903
10	-0.083	1.235	0.115	0.686

$\bar{b}_1 = -1.513 \quad \bar{b}_2 = -0.985$

$\hat{\sigma}_1^2 = 1.343 \quad \hat{\sigma}_2^2 = 1.055$

$\hat{\rho}_{12} = 0.769$

แทนค่าในสมการ 2b จะได้

$\hat{m} = -0.985 - (0.798)(-1.513) = 0.222$

แทนค่า $\hat{k}$ และ $\hat{m}$ ในสมการ 1a และ 1b จะได้

$\hat{b}_{1g} = 0.798\,\hat{b}^*_{1g} + 0.222$

และ

$$\hat{a}_{1g} = \frac{\hat{a}^*_{1g}}{0.798}$$

เท่ากับสมการสำหรับการแปลงรูปจากสเกลเพศชายไปเป็นสเกลเพศหญิง

ตาราง 8 รายงานค่าของ $\hat{b}_{1g}$ และ $\hat{a}_{1g}$ ที่อ้างอิงจากสมการ 1a และ 1b รายงานค่า $\hat{b}_{2g}$, $\hat{a}_{2g}$ และ A_g ในการวัดพื้นที่ของรันเนอร์ (Rudner) สำหรับข้อสอบที่ลำเอียง ข้อที่ 1 จะมี

ค่าพื้นที่มากที่สุดสำหรับข้อสอบที่ลำเอียง และข้อ 8 จะมีค่าการวัดพื้นที่ที่น้อยที่สุด ใช้ Rules of thumb สำหรับการตัดสินใจเมื่อข้อสอบลำเอียงจะไม่ถูกต้องเสมอไป ในทางปฏิบัติโดยทั่วไปจะดูข้อสอบที่สงสัยว่าลำเอียงโดยการพิจารณาพื้นที่ในการวัดจะมีมาก ข้อสอบที่สงสัยว่าจะมีความลำเอียงคือข้อที่ 1, 5 และอาจจะมีข้อที่ 3 ด้วย เมื่อข้อเหล่านี้มีพื้นที่การวัดความลำเอียงที่มาก

ตาราง 8 พื้นที่การวัด (A_g) ของความลำเอียง

ข้อสอบ	ชาย		หญิง		
	$\hat{b}_{1g}$	$\hat{a}_{1g}$	$\hat{b}_{2g}$	$\hat{a}_{2g}$	$\hat{A}_g$
1	−2.753	0.672	−1.467	1.198	1.13
2	0.130	1.252	0.222	0.723	0.47
3	−0.959	3.167	−1.602	0.989	0.77
4	−2.146	0.531	−1.734	0.592	0.32
5	−1.865	0.715	−3.202	0.589	1.02
6	−0.875	2.167	−1.278	0.909	0.60
7	−0.283	1.103	0.139	0.718	0.52
8	−0.733	1.162	−0.428	0.963	0.31
9	−0.531	2.405	−0.624	0.903	0.55
10	0.155	1.546	0.115	0.686	0.64

อย่างไรก็ตามในการใช้เทคนิคคุณลักษณะแฝงจะมี 2 วัตถุประสงค์ของงานวิจัยที่ศึกษาข้อสอบที่ลำเอียง นั่นคือ เหตุผลแรก ถ้า ICC ทั้งหมดเท่ากันในทุก ๆ กลุ่มประชากรย่อย ชุดของข้อสอบจะไม่ลำเอียงตามนิยามของทฤษฎีคุณลักษณะแฝง และไม่ลำเอียงตามนิยามอื่น ๆ อย่างไรก็ตาม ข้อสอบซึ่งลำเอียงตามนิยามของคุณลักษณะแฝง ไม่จำเป็นต้องลำเอียงตามนิยามอื่น ๆ

เหตุผลที่สอง ICC ที่เปลี่ยนไปในกรณีที่ข้อสอบวัดคุณลักษณะแฝงมิติเดียวถูกวัดในกลุ่ม 2 กลุ่ม ในสถานการณ์นี้บ่งชี้ถึงข้อสอบที่ลำเอียง เหตุผลในการแปรเปลี่ยนของ ICC ที่พบอาจจะเปิดเผยโดยการแสดงความเป็นมิติของข้อสอบในแต่ละกลุ่ม ถ้าข้อสอบเป็นมิติเดียวในแต่ละกลุ่ม การแปรเปลี่ยนของ ICC จะเนื่องมาจากการวัดคุณลักษณะที่แตกต่างกันในแต่ละกลุ่ม แต่ถ้าข้อสอบมีหลายมิติในแต่ละกลุ่ม ความแปรเปลี่ยนของ ICC จะเนื่องมาจากมิติที่หลากหลายของข้อสอบ

ข้อสอบที่เป็นมิติเดียวในแต่ละกลุ่ม แต่มี ICC ไม่เหมือนกัน ดังนั้นจึงมีความเหมาะสมที่จะทิ้งข้อสอบที่มี ICC ที่แตกต่างกันและอาจกล่าวได้ว่า ข้อสอบที่เหลือนั้นไม่ลำเอียง เพราะว่าชุดของข้อสอบดั้งเดิมเป็นมิติเดียวสำหรับในแต่ละกลุ่ม และถ้าหากข้อสอบที่เหลือในแบบทดสอบนั้นวัดในคุณลักษณะที่แตกต่างกันในทั้งสองกลุ่ม ในทางปฏิบัติเป็นไปได้ที่ข้อสอบนั้นวัดในหลาย

มิติ ดังนั้นถ้ามีข้อสอบเพียงไม่กี่ข้อที่เปลี่ยนแปลงเมื่อใช้กับกลุ่มที่แตกต่างกัน และข้อสอบมีดัชนีอำนาจจำแนกต่ำ จึงมีเหตุผลเพียงพอที่จะทิ้งข้อสอบนั้นและข้อสอบที่เหลืออยู่จะไม่ลำเอียง

บรรณานุกรม

Camilli, Gregory and Shepard, Lorrie A. **Methods of Identifying Biased Test Items**. Thousand Oaks : SAGE Publications, Inc., 1994.

Crocker, Linda and Algina, James. **Introduction to Classical and Modern Test Theory**. New York : CBS College Publishing, 1986.

Ostelind, Steven J. **Test Item Bias**. Beverly Hills : SAGE Publications, Inc., 1983.

จัดทำเสร็จเดือนมกราคม พ.ศ.2547