4. ตรวจสอบโมเดลดัชนีปรับแก้ (Modification Indices)

5. เพิ่มอิทธิพลตรงจากความแปรปรวนร่วมกับข้อคำถามที่มีค่าดัชนีปรับแก้สูงสุด

6. ดำเนินในขั้นตอนที่ 4 – 5 จนกว่าจะไม่พบค่าดัชนีปรับแก้ (M.I.) ที่ไม่มีนัยสำคัญ ประเมินความสอดคล้องของโมเดลและอิทธิพลตรง (β_i)

ตอนที่ 5 แนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบด้วยวิธี BAYESIAN

ทฤษฎีของ Bayesian

เดิมวิธีการของ Classic ได้กำหนดให้ค่า θ คงที่ตามจำนวนตัวอย่างที่ถูกเก็บมา แต่ในขณะที่ Bayesian Approach ได้ให้ค่า θ เป็นค่าของ Random Variable (Θ)

สำหรับแนวความคิดแบบเบส์ ตัวแบบของค่าสังเกต $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)^T$ ที่กำหนดเงื่อนไขบนค่าพารามิเตอร์ที่ไม่ทราบค่า $\theta = (\theta_1, \theta_2, ..., \theta_p)^T$ จะอยู่ในรูปของการแจกแจงความน่าจะเป็น $\int x|\theta$ ซึ่งมาจากการที่เรากำหนดให้ θ เป็นตัวแปรสุ่ม Random Variable (Θ)

การที่ θ เป็นการแจกแจงก่อน (Prior) นั้นเราจะให้สัญลักษณ์ของ Prior Density ของ θ เป็น $\pi(\theta)$ นั่นคือ "Prior" จะถูกใช้แทนค่าของ Density θ ดังนั้นถ้าเราเก็บข้อมูล X = x เราจะได้เงื่อนไขของ Density Given X= x ซึ่งเรียกว่า "Posterior" เขียนได้เป็น $x|\theta$ และมีการแจกแจงก่อน (Prior Distribution) เป็น $\pi(\theta|\eta)$ โดยที่ η คือ เวกเตอร์ของพารามิเตอร์ขั้นที่สอง (Hyperparameter) ในแนวคิดแบบเบส์นี้การอนุมานเกี่ยวกับ θ จะขึ้นอยู่กับการแจกแจงภายหลัง (Posterior Distribution) ซึ่งมีรูปแบบเป็น

$$p(\theta|x,\eta) = \frac{p(x,\theta|\eta)}{p(x|\eta)} = \frac{\int(x|\theta)\pi(\theta|\eta)}{p(x|\eta)}$$

เมื่อ
$$p(x|\eta) = \begin{cases} \sum_{\theta} \int(x|\theta)\pi(\theta|\eta) & ,\theta \text{ is discreate} \\ \int f(x|\theta)\pi(\theta|\eta)d\theta & ,\theta \text{ is continuous} \end{cases}$$

โดยสมการดังกล่าว ถูกเรียกว่า ทฤษฎีของเบส์ (Bayes' Theorem) จะสามารถสังเกตได้ว่าข้อมูลจะมีอยู่สองช่วงนั่นคือ ข้อมูลที่ได้จากการทดลองหรือข้อมูลใหม่ที่ได้รับ (ซึ่งอยู่ในรูปของความน่าจะเป็น *f*) และความเชื่อก่อน (ซึ่งอยู่ในรูปของความน่าจะเป็นก่อนการทดลอง π) โดยค่าของอินทิกรัลของตัวส่วนในสมการ จะถูกเรียกว่า การแจกแจงหน่วยสุดท้าย (Marginal Distribution) ของ *x* เมื่อกำหนดพารามิเตอร์ขั้นที่ 2 เป็น η และเขียนได้ในรูปของ $m(x|\eta)$

อย่างไรก็ตามหากเราสามารถทราบค่าของ η เราสามารถที่จะตัดตัวแปรนี้ทิ้งออกจากระบบสมการได้ เนื่องจากไม่มีความจำเป็นที่จะต้องกำหนดเงื่อนไขบนค่าคงที่ ดังนั้นเราสามารถที่จะเขียนรูปแบบของการแจกแจงใหม่ โดยให้อยู่ในรูปแบบที่ง่ายขึ้นเป็น

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)} = \frac{\int(x|\theta)\pi(\theta)}{p(x)}$$

เมื่อ $$p(x|\eta) = \begin{cases} \sum_{\theta} \int(x|\theta)\pi(\theta|\eta) & ,\theta \text{ is discreate} \\ \int f(x|\theta)\pi(\theta|\eta)d\theta & ,\theta \text{ is continuous} \end{cases}$$

ถ้าเราทำการพิจารณา ค่าพารามิเตอร์ที่มีลักษณะต่อเนื่อง จากทฤษฎีของเบส์ เราสามารถสร้างการกระจายของ $\theta|x$ ที่เรียกว่า Posterior distribution ของ θ โดยเราจะให้ Posterior Distribution ของ θ เป็น $p(\theta|x)$ ซึ่งสามารถเขียนได้เป็น

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}$$

นั่นก็คือ $L(x|\theta)$ หรือ $f(x|\theta)$ เป็น Likelihood function และสัญลักษณ์ Θ แทนค่าเป็น Parameter Space ของ θ หรือ Support of $\pi(\theta)$ ซึ่งมีค่าเป็น

$$p(x) = \int_{\Theta} L(x|\theta)\pi(\theta)d\theta$$

ซึ่งเป็นค่า Normalizing Constant ของ Posterior Distribution ของ θ โดยที่ $p(x)$ เป็น Marginal Probability Distribution ของ x สำหรับการอนุมานปัญหาของ $p(x)$ จะไม่มีรูปแบบที่แน่นอนของการอนุมานแบบ Bayesian ของค่า θ ที่เป็นพื้นฐานของการแจกแจงภายหลังของ θ, $p(\theta|x)$ โดยตัว Posterior เอง แล้วสามารถคำนวณหาค่าข้อมูลได้หลายอย่าง เช่น การหาค่าเฉลี่ย, ค่ามัธยฐาน, ฐานนิยม, ความแปรปรวนและควอไทล์

การอนุมานแบบ Bayesian โดยสรุปแล้วเราจะสามารถหาค่าได้จากการให้ข้อมูลดังนี้

ข้อมูล	$x = (x_1, x_2, \ldots, x_n)^T$
ค่าพารามิเตอร์	$\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$
Likelihood	$L(x\|\theta)$
Prior	$\pi(\theta)$

โดยการอนุมานอยู่บนพื้นฐานของ Joint Posterior ดังนั้น

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}$$

จากสมการ (3) ถ้าเราไม่สนใจว่าค่าของ ที่เป็นค่าที่ไม่ขึ้นอยู่กับ θ โดยคิดว่า x เป็นค่าคงที่ เราสามารถที่จะทำสมการ (3) ในรูปแบบที่ง่ายขึ้นอีก นั่นคือ

$$p(\theta|x) \propto f(x|\theta)\pi(\theta)$$

Posterior $\propto$ *Likelihood* $\times$ *prior*

ความน่าจะเป็นภายหลังจะเป็นสัดส่วนกับความควรจะเป็นคูณด้วยความน่าจะเป็นก่อนหน้า

ถ้าเราไม่แน่ใจความเหมาะสมของค่า η เราสามารถกำหนดให้ความไม่แน่นอนนี้อยู่ในรูปของการแจกแจงก่อนขั้นที่ 2 (Second – stage Prior Distribution หรือ Hyperparameter) โดยสามารถเขียนแทนสมการดังกล่าวด้วย $h(\eta)$ ซึ่งมีการแจกแจงภายหลังของค่า θ เป็น

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)} = \frac{\int f(x,\theta,\eta)d\eta}{\int\int f(x,\theta,\eta)d\eta d\theta}$$

$$= \frac{\int f(x|\theta)\pi(\theta|\eta)h(\eta)d\eta}{\int\int f(x|\theta)\pi(\theta|\eta)h(\eta)d\eta d\theta}$$

ถ้าเราสมมติให้เหตุการณ์มี 2 เหตุการณ์ คือ A และ B จะได้เงื่อนไขความน่าจะเป็นก็จะเป็นดังนี้

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{Pr[B|A]Pr[A]}{Pr[B]}$$

Markov Chain Monte Carlo

เป็นวิธีการของ Computationally Intensive Simulation เพื่อแทนการทำวิธี Integrals ซึ่งถูกพัฒนาในปี ค.ศ. 1980 โดยทำการสร้างโอกาสความเป็นไปได้ในการรับมือกับปัญหาที่แท้จริงที่เกิดขึ้นในปัจจุบันที่มีความซับซ้อนมากยิ่งขึ้น

วิธีการจำลองของ Markov Chain คือ การสร้าง Markov Process ซึ่งมี Stationary Transition Distribution ให้กลายเป็นสิ่งที่สามารถระบุได้ นั่นคือ $p(\theta|x)$ และทำการจำลองส่วนที่จำเป็นในระยะยาว ดังนั้น การกระจายของค่าปัจจุบันของกระบวนการนี้จะมีค่าที่แน่นอนพอที่จะนำไปสู่ Stationary Transition Distribution

ดังนั้นเราสามารถอ้างวิธีการใช้ Markov Chain Simulation เพื่อที่จะได้ Distribution ของ $p(\theta|x)$ ซึ่งจะถูกเรียกว่า Markov Chain Monte Carlo (MCMC) Method

Gibbs Sampler

Gibbs Sampler เป็นส่วนหนึ่งของ MCMC Class ซึ่งง่ายที่จะทำการอธิบายและเข้าใจโดยทำการยกตัวอย่าง ดังนี้

กำหนดให้ $\theta = [\theta_1\ \theta_2]$ ซึ่งมี Posterior Density $p(\theta) = p(\theta_1, \theta_2)$ โดยที่รู้อย่างง่าย ๆ คือ เราสามารถที่จะไม่ต้องใช้ตัวแปรตามของ y ได้ ถ้าเรารู้เงื่อนไข Conditional Density ซึ่งไม่ต้องรับรองในเงื่อนไขมากนักก็ได้ แต่เรารู้ได้จาก $p(\theta_1|\theta_2)$ และ $p(\theta_2|\theta_1)$ ซึ่งจำเป็นต้องมีน้อยกว่าทางเลือก Sequential Draws จาก $p(\theta_1|\theta_2)$ และ $p(\theta_2|\theta_1)$ ในการจำกัดจากการเลือกจาก $p(\theta_1|\theta_2)$ เช่นถ้าเราพิจารณากรณีอย่างง่าย คือมีค่าพารามิเตอร์ 3 ตัว $(\theta_1, \theta_2, \theta_3)$ แทนเป็น 3 เงื่อนไขของ Posterior Distribution เมื่อทำการหา Iterative 1 รอบ จะได้จาก

$$f_1(\theta_1|\theta_2\theta_3, y)$$
$$f_2(\theta_2|\theta_3\theta_1, y)$$
$$f_3(\theta_3|\theta_1\theta_2, y) \quad \text{โดยที่ } y = (y_1, y_2, \ldots, y_n)^T$$

ซึ่งกระบวนการของ Gibbs Sampling เป็นดังนี้

1. พิจารณา Arbitrary Set ของ Starting Parameter Value ซึ่งจะได้ค่า $\theta_{1,0}, \theta_{2,0}, \theta_{3,0}$

2. สร้าง M + N Set ของ Random Number โดยทำการ Iterative จาก Full Conditional Posterior Distribution เราจะได้อันดับแบบ General Form แบบ *i – th* เป็น $\{\theta_{1,i}, \theta_{2,i}, \theta_{3,i}\}$ ดังนั้นเราจะได้

2.1 $\theta_{1,i+1}$ จาก $f_1\left(\theta_1|\theta_{2,i},\theta_{3,i},y\right)$

2.2 $\theta_{2,i+1}$ จาก $f_2\left(\theta_2|\theta_{3,i},\theta_{1,i+1},y\right)$ และ

2.3 $\theta_{3,i+1}$ จาก $f_3\left(\theta_3|\theta_{1,i+1},\theta_{2,i+1},y\right)$ จากอันดับ (i +1) – the Realization

3. ทิ้งตัวแรก M ที่ได้จากการรวมในขั้นที่ 2 แล้วใช้ค่า N ที่เป็นตัวสุดท้ายเพื่อทำการสร้างแบบของ Random Sample $\{(\theta_{1,i},\theta_{2,i},\theta_{3,i})\}_{i=M+1}^{M+N}$ และทำการประมาณค่า Posterior Marginal โดยใช้ Random Sample

ภายใต้เงื่อนไขอย่างง่าย Geman & Geman (1984) ได้แสดงว่า Joint Distribution ของ Above Random Sample จะ Convergence Exponentially สู่ Joint Posterior Distribution ของ $(\theta_1,\theta_2,\theta_3)$ ดังนั้นก่อนที่จะ Realization Form ของ Random Sample จาก Joint Distribution $f_1(\theta_1,\theta_2,\theta_3|y)$ เราสามารถทำการอนุมานได้จาก

$$\hat{\theta}_i = \frac{1}{N \quad M} \sum_{j=M+1}^{N} \theta_{i,j}$$

$$\hat{\sigma}^2{}_i = \frac{1}{N \quad M \quad 1} \sum_{j=M+1}^{N} (\theta_{i,j} \quad \hat{\theta}_i)^2$$

โดยค่า Credible Interval $\{l,\mu\}$ ของ θ_i อยู่ภายใต้เงื่อนไข $p(l\ \theta_i\ \mu)= 0.95$ เป็นต้น เป็น Algorithm อีกแบบซึ่งเป็นการแทนที่ค่าที่เป็นไปได้ในแต่ละ State เพื่อทำการหาค่าที่ดีที่สุด ในที่นี้เป็นค่าของความเสี่ยงและ Probability Distribution Function (PDF) เพื่อเป็นการตัดสินว่าค่าที่ได้เป็นค่าที่ดีที่สุด นั่นคือ เป็นหนึ่งในวิธีการหาค่าที่ดีที่สุดของข้อมูล (Finding Best Solution)

Bayes Factor

ในทางสถิติ จะใช้ Bayes Factors เป็นทางเลือก Bayesian ที่จะทำการทดสอบ Classical Hypothesis นั่นคือ การกำหนดข้อมูลภายใต้สมมติฐานหนึ่งหนึ่ง โดยใช้ปัญหาของการเลือกแบบจำลองซึ่งทำการเลือกระหว่างแบบจำลอง M_i และ M_j บนพื้นฐานข้อมูล Vector y โดย M_i และ M_j เปรียบเหมือนข้อสมมติฐานที่ตั้งขึ้น แต่ละแบบจำลองก็จะมีค่าความน่าจะเป็นของตัวเอง เช่น $p(y|M_i)$ หรือ $p(y|M_j)$ และกำหนดให้ความน่าจะเป็นก่อน (Prior Probability) เป็น $p(M_i)$และ $p(M_i)= 1 - p(M_j)$ แล้วจะได้ความน่าจะเป็นภายหลัง (Posterior Probability) เป็น $p(M_i|y)$ และ $p(M_i|y)= 1 - p(M_j|y)$ เนื่องจากความเชื่อก่อนจะแปลงข้อมูลที่ได้ไปเป็น

ความเชื่อภายหลังเมื่อนำข้อมูลที่มีอยู่เข้ามาทำการพิจารณา ซึ่งเราสามารถที่จะทำการแปลงข้อมูลเพื่อให้ได้ในรูปของอัตราส่วนออดส์ได้

$$\text{Odds} = \frac{\text{probability}}{\text{1- probability}}$$

ดังนั้น Bayes Factor: BF จะได้เป็น

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)}$$

จะกล่าวได้ว่า

Posterior odds = Bayes factor × prior odds

ปัจจัยเบส์เป็นอัตราส่วนของ Posterior odds ของ ต่อ prior odds *iM* โดยที่

$$p(y|M_i) = \int p(x|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i$$

โดยที่ θ_i คือ เวกเตอร์ของพารามิเตอร์ภายใต้ M_i

$\pi(\theta_i|M_i)$ คือ การแจกแจงก่อนของ θ_i

$p(y|\theta_i, M_i)$ คือ ความน่าจะเป็นของ *y* เมื่อกำหนด θ_i หรือความควรจะเป็นของ θ

ดังนั้น ถูกเรียกว่า Marginal Likelihood สำหรับ Model *i* โดยทั่วไป แบบจำลอง M_i และ M_j จะเป็น Parametrised โดย Vector ของ Parameters θ_i และ θ_j ดังนั้น BF จะเป็น

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)} = \frac{\int p(\theta_i|M_i)p(y|\theta_i, M_i)d\theta_i}{\int p(\theta_j|M_j)p(y|\theta_j, M_j)d\theta_j}$$

ถ้าให้ $p(M_i)$ คือ Prior Probability ของแบบจำลอง M_i

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)} \div \frac{p(M_i)}{p(M_j)}$$

สมการนี้คือ Posterior Odds ใน Favor ของแบบจำลอง M_i ซึ่งถูกแยกโดย Posterior Odds ใน Favor ของแบบจำลอง M_j

ในทฤษฎีความน่าจะเป็น สถิติ การอนุมาน และปัญญาประดิษฐ์บางครั้งจะพบคำว่า แบบเบส์ (Bayesian) มาขยายชื่อทฤษฎีหรือโมเดลต่าง ๆ โดยทุกครั้งที่พบคำขยายนี้หมายความว่า ได้มีการนำปรัชญาหรือหลักการของ ทฤษฎีความน่าจะเป็นแบบเบส์ (บางท่านเรียกการอนุมานแบบเบส์ หรือ สถิติแบบเบส์) มาใช้กับสาขาความรู้นั้น ๆ

ถ้าจะกล่าวอย่างไม่เป็นทางการทฤษฎีความน่าจะเป็นแบบเบส์แปลความหมายของคำว่า ความน่าจะเป็น เป็นความเชื่อมั่นส่วนบุคคลในเหตุการณ์หนึ่ง ๆ ซึ่งต่างจากทฤษฎีความน่าจะเป็นของคอลโมโกรอฟ (ที่มักถูกเรียกว่าทฤษฎีความน่าจะเป็นเชิงความถี่) ที่มักแปลความหมายของความน่าจะเป็น (โดยต้องแปลควบคู่ไปกับการทดลองเสมอ) ดังนี้ความน่าจะเป็นของเหตุการณ์ A คือ อัตราส่วนของจำนวนครั้งของเหตุการณ์ A ที่ทดลองสำเร็จเทียบกับจำนวนครั้งที่ทดลองทั้งหมด จุดแตกต่างสำคัญระหว่างทฤษฎีทั้งสองประเภทมีดังนี้

1. ความหมายของความน่าจะเป็น พวกเบส์มองความน่าจะเป็น เป็นความเชื่อส่วนบุคคล พวกเชิงความถี่มองความน่าจะเป็น เป็นคุณสมบัติหนึ่งที่ถูกฝังอยู่ในวัตถุ (ไม่ขึ้นกับตัวบุคคล)

2. การนำทฤษฎีไปใช้งาน ในการนำทฤษฎีความน่าจะเป็นเชิงความถี่ไปใช้จะต้องมีการทดลองเชิงแนวคิด (Conceptual Experiment) ควบคู่ไปด้วยเสมอ เหตุการณ์ใด ๆ ก็ตามที่ไม่มีการทดลองเชิงแนวคิดที่สมเหตุสมผลพอจะไม่สามารถนำทฤษฎีความน่าจะเป็นเชิงความถี่ไปใช้งานได้ เช่น ไม่สามารถจินตนาการการทดลองเพื่อทดสอบว่ามีมนุษย์ต่างดาวอยู่หรือไม่ได้ ฉะนั้นประโยค ความน่าจะเป็นที่จะมีมนุษย์ต่างดาวไม่มีความหมายในทฤษฎีความน่าจะเป็นเชิงความถี่ แต่สามารถนำทฤษฎีความน่าจะเป็นแบบเบส์มาอ้างความน่าจะเป็นประเภทนี้ได้ ในมุมมองนี้อาจกล่าวได้ว่าทฤษฎีความน่าจะเป็นแบบเบส์สามารถนำไปประยุกต์ใช้งานได้กว้างขวางมากกว่า

กล่าวโดยสรุปทฤษฎีความน่าจะเป็นแบบเบส์ มีปรัชญาที่ต่างจากทฤษฎีความน่าจะเป็นเชิงความถี่เกือบสิ้นเชิงถึงแม้จะมีสัจพจน์พื้นฐานแบบเดียวกัน โดยในทฤษฎีความน่าจะเป็นแบบเบส์นั้นมองความน่าจะเป็นสถิติหรือการอนุมานเป็นเรื่องเดียวกัน (ชัยวัช โชวเจริญสุข, 2552)

วิธีของเบส์ (Bayesian Estimation)

เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด มีข้อจำกัดและปัญหาเมื่องทำการประมาณค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้เข้าสอบพร้อม ๆ กัน วิธีของเบส์จึงอาจเป็นวิธีที่เหมาะสมกว่าในการประมาณค่าพารามิเตอร์ (วิชุดา บัวคง, 2533, หน้า 47) ซึ่งแนวคิดของเบส์มีแนวคิดบางประการต่างออกไป จากแนวคิดของวิธีแมกซิมัมไลค์ลิฮูด และได้รับการพัฒนาให้สามารถประมาณค่า θ และ a, b, c ร่วมกันได้อย่างมีประสิทธิภาพ (Swaminathan & Gifford, 1985) จากการนิยามความน่าจะเป็นของผู้เข้าสอบในการตอบข้อสอบข้อที่ j เมื่อ U_j=1 สำหรับการตอบถูก และ U_j = 0 สำหรับการตอบผิด แสดงได้ดังสมการ

$$P(U/\theta,b,a,c) = P(U_i = 1/\theta,b,a,c) \bullet P(U_j = 0/\theta,b,a,c)$$
$$= P_j^{U_j} . Q_j^{1-U_j} \bullet Q_j = 1 - P_j$$

ถ้าผู้เข้าสอบตอบข้อสอบ n ข้อ และข้อสอบมีความเป็นอิสระเฉพาะที่แล้ว ความน่าจะเป็นของการตอบ แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม ดังสมการต่อไปนี้

$$P(U_1,U_2,...,U_n/\theta,b,a,c) = \prod_{j=1}^{n} P_j^{U_j} Q_j^{1-U_j}$$

จากสมาการดังกล่าวข้างต้น เป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อ ที่สามารถจัดหรือสังเกตได้โดยที่ $U_1, U_2,...,U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $u_1, u_2,...,u_n$ เมื่อ u_j มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ, b, a, c ที่จะบอกว่าตัวแปรสุ่มนี้ มีโอกาสเกิดขึ้นเพียงใด จึงเรียกฟังก์ชันนี้ว่า ฟังก์ชันความน่าจะเป็นหรือฟังก์ชันไลค์ลิฮูด (Likelihood Function) ซึ่งแสดงได้ดังนี้

$$L(u_1,u_2,...,u_n/\theta,b,a,c) = \prod_{j=1}^{n} P_j^{U_j} Q_j^{1-U_j}$$

เมื่อ u_j = 1 ค่าของ Q_j จะหมดไป และเมื่อ u_j = 0 ค่าของ P_j ก็จะหมดไป และฟังก์ชันไลค์ลิฮูด ที่มีผู้เข้าสอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(u/\theta,b,a,c) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_j^{U_j} Q_j^{1-U_j}$$

เมื่อ u = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

$$P_{ij} = P_j(\theta_j, b_j, a_j, c_j)$$

การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ มีแนวคิดว่า ค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c เป็นตัวแปรสุ่ม (Random Variable) จากการแจกแจงที่แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม (Joint Density Function) $f(\theta, b, a, c)$ และเรียกฟังก์ชัน $f(\theta, b, a, c)$ นี้ว่า การแจกแจงเริ่มแรก (Prior Distribution) ของค่า θ, b, a และ c ซึ่งทำให้การใช้ $L(U/\theta, b, a, c)$ เพียงอย่างเดียวในการประมาณค่า θ, b, a และ c ถูกพิจารณาว่าเป็นการใช้ข้อมูลที่มีอยู่อย่างไม่ครบถ้วน เพราะยังมีการแจกแจงเริ่มต้นร่วมกับ $f(\theta, b, a, c)$ ที่ควรนำมาใช้ในการประมาณค่าพารามิเตอร์ด้วย กล่าวคือถ้าพิจารณาความน่าจะเป็นร่วมของการตอบข้อสอบ $P(U/\theta, b, a, c)$ จะเห็นว่าการแจกแจงของตัวแปร U ขึ้นอยู่กับค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c ถ้าค่า θ, a, b และ c เปลี่ยนไป โอกาสที่ตัวแปร U มีค่าเท่ากับ u ก็จะเปลี่ยนไปด้วย ดังนั้นการทราบผลการตอบข้อสอบ u จึงน่าจะช่วยทำให้ทราบค่า θ, a, b และ c ได้ดียิ่งขึ้น ซึ่งอาจแสดงได้ด้วยการแจกแจงอย่างมีเงื่อนไขของค่า θ, a, b และ c เมื่อทราบผลการตอบข้อสอบ $f(\theta, b, a, c/ U)$ และเรียกว่า การแจกแจงภายหลัง (Posterior Distribution)

การแจกแจงภายหลังร่วมกันของค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c เมื่อทราบผลการตอบข้อสอบ u คือ

$$f(\theta, b, a, c/ U) = L(u/\theta, b, a, c)\, f(\theta, b, a, c)/ f(u)$$

เมื่อ $f(u)$ = การแจกแจงมาร์จินัล(Marginal)ของผลการตอบสนองข้อสอบ

$f(\theta, b, a, c)$ = การแจงแจงเริ่มแรกของค่า θ, a, b และ c

$f(U/\theta, b, a, c)$ = ฟังก์ชันไลค์ลิฮูด

$f(\theta, b, a, c/ U)$ = การแจกแจงภายหลังของค่า θ, a, b และ c เมื่อทราบผลการตอบสนองต่อข้อสอบ u

ซึ่งการแจกแจงภายหลังเป็นฟังก์ชันที่ใช้ในการประมาณค่า θ, a, b และ c ด้วยวิธีของเบส์โดยมีส่วนที่แตกต่างจากการประมาณด้วยวิธีแมกซิมัมไลค์ลิฮูด คือ การแจกแจงเริ่มแรก $f(\theta, b, a, c)$ และการแจกแจงมาร์จินัล $f(u)$ ซึ่งการแจกแจงมาร์จินัลนี้ เป็นการแจกแจงที่ไม่ขึ้นอยู่กับค่า θ, a, b และ c จึงถือว่าเป็นค่าคงที่ในการประมาณค่า θ, a, b และ c

เนื่องจากแนวคิดจากการประมาณค่าด้วยวิธีของเบส์ ดังกล่าว ทำให้สามารถจำแนกกระบวนการดำเนินการตามแนวคิด ออกได้เป็น 2 กระบวนการ ดังนี้

1. กระบวนการกำหนดลักษณะของการแจกแจงเริ่มแรก มี 2 ขั้น (Swaminathan & Gifford, 1985, pp. 589 – 601) คือ

1.1 กำหนดให้การแจกแจงเริ่มแรกของค่าความสามารถ (θ) ค่าอำนาจจำแนก (*a*) ค่าความยาก (*b*) และค่าการเดา (*c*) เป็นอิสระต่อกัน

$$f(\theta,b,a,c) = f(\theta) \cdot f(b) \cdot f(a) \cdot f(c)$$

และกำหนดลักษณะของการแจกแจงของ $f(\theta)$. *f(b)* . *f(a)* และ *f(c)* ดังนี้

1.1.1 การแจกแจงเริ่มแรกของค่าความสามารถ $f(\theta)$ มีข้อตกลงว่าข้อสารสนเทศที่มีมาก่อนของค่าความสามารถของผู้เข้าสอบแต่ละคนไม่แตกต่างกัน สามารถใช้แทนกันได้ (Exchangeability) และค่าความสามารถเป็นตัวแปรสุ่มที่มีการแจกแจงเป็นปกติ (Normal Distribution)

$$f(\theta_i / \mu_\theta, \sigma_\theta^2) = N(\mu_\theta, \sigma_\theta^2)$$

เมื่อ $N(\mu_\theta, \sigma_\theta^2)$ = การแจกแจงปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเท่ากับ μ_θ และค่าความแปรปรวนเท่ากับ σ_θ^2

1.1.2 การแจกแจงเริ่มแรกของค่าความยากของข้อสอบ *f(b)* อาจใช้กระบวนการเดียวกับการกำหนดการแจกแจงเริ่มแรกของค่าความสามารถ คือ มีข้อตกลงว่า *f(b)* มีการแจกแจงเป็นปกติ หรืออาจจะไม่กำหนดการแจกแจงเริ่มแรกไว้ก็ได้ (Swaminathan & Gifford, 1985, p. 350)

1.1.3 การแจกแจงเริ่มแรกของค่าอำนาจจำแนกของข้อสอบ *f(a)* เนื่องจากค่าอำนาจจำแนกของข้อสอบโดยทั่วไปจะต้องเป็นค่าบวก และเป็นความชันของเส้นโค้งลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง ดังนั้น การแจกแจงเริ่มแรกของค่าอำนาจจำแนก a_j จึงควรเป็นการ แจกแจงแบบไคว์ (Chi Distribution)

$$f(a_j / v_j w_j) \alpha a_j^{v_j - 1} exp[a_j^2 / 2w_j]$$

เมื่อ v_j = Degree of Freedom

w_j = Scale Parameter

1.1.4 การแจกแจงเริ่มแรกของค่าการเดาของข้อสอบ $f(c)$ เนื่องจากค่าพารามิเตอร์ c_j จะมีขอบเขตอยู่ตั้งแต่ 0 – 1 ดังนั้น จึงมีข้อตกลงว่าควรมีการแจกแจงเบต้า (Beta Distribution)

$$f(c_j/s_j,t_j) \propto c_j^{s_j}(1-a_j^2)$$

เมื่อ s_j,t_j = Scale Parameter

1.2 กำหนดค่าที่เป็นตัวเลขของพารามิเตอร์ สำหรับการแจกแจงเริ่มแรก

1.2.1 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ได้แก่ μ_θ และ σ_θ^2 อาจกำหนดให้ $\mu_\theta = 0$ และ $\sigma_\theta^2 = 1$ ซึ่งจะทำให้การประมาณค่าความสามารถ θ มีความสะดวก และประมาณค่าได้รวดเร็วขึ้น (Swaminathan & Gifford, 1985, pp. 351 – 355)

1.2.2 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความยาก b หากมีการกำหนดลักษณะการแจกแจงไว้ ก็ใช้ค่าเดียวกับพารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ดังกล่าวแล้วข้างต้น

1.2.3 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าอำนาจจำแนก a ได้แก่ v_j, w_j การกำหนดค่าของ v_j และ w_j ที่เหมาะสม อาจจะหาได้จากการกำหนดพิสัย (Range) ของค่า a คือ ถ้าให้ H เป็นขีดจำกัดบนของพิสัย และให้ L เป็นขีดจำกัดล่างของพิสัย จะหาค่า v_j และ w_j ได้จากสูตร

$$v_j = \frac{1}{2}(1 + Z_{\frac{1}{2}}((H+L)/(H-L))^2)$$

$$w_j = \frac{1}{2}((H-L)/Z_{(1/2)\alpha})^2$$

เมื่อ $Z_{(1/2)\alpha}$ = ค่า Z ของการแจกแจงปกติมาตรฐาน (Standard Normal Distribution) ที่ระดับนัยสำคัญ α

v_j = Degree of Freedom

นอกจากวิธีดังกล่าวอาจจะกำหนดให้ v_j และ w_j ของการแจกแจงเริ่มแรก $f(a)$ ของข้อสอบทุกข้อเท่ากัน คือ $v_j = 10$ และ $w_j = 0.1$ ซึ่งจะทำให้การประมาณค่าพารามิเตอร์มีความสะดวกยิ่งขึ้น (Swaminathan & Gifford, 1985, pp. 356 – 357) และการกำหนดเช่นนี้ก็ยังคงทำให้ค่าประมาณของพารามิเตอร์ a_j ตกอยู่ในช่วงที่ยอมรับได้ คือ $0.40 < a_j < 1.55$ ด้วยความเชื่อมั่น 99 % (Swaminathan & Gifford, 1985, p. 356)

1.2.4 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าการเดา c ได้แก่ s_j และ t_j การกำหนดค่าของ s_j และ t_j ที่เหมาะสม อาจจะหาได้จากการสังเกตสัดส่วนการตอบถูกของกลุ่มผู้เข้าสอบที่มีความสามารถในระดับต่ำมาก กล่าวคือ ถ้าให้ m แทนจำนวนผู้เข้าสอบที่มีความสามารถระดับต่ำมาก และให้ M แทนสัดส่วนการตอบถูกของกลุ่ม m แล้วจะหาค่า s_j และ t_j ได้จากสูตร

$$s_j = mM \qquad \text{และ} \qquad t_j = m(1 - M) - 2$$

นอกจากวิธีนี้ อาจจะกำหนดให้ s_j และ t_j ของการแจกแจงเริ่มแรก $f(c)$ ของข้อสอบทุกข้อเท่ากัน คือ $s_j = 2$ และ $t_j = 12$ ซึ่งจะทำให้ประมาณค่าพารามิเตอร์มีความสะดวกยิ่งขึ้น (Swaminathan & Gifford, 1985, pp. 589 – 601) และการกำหนดเช่นนี้ก็ยังคงทำให้ค่าประมาณของพารามิเตอร์ c_j ตกอยู่ในช่วงที่ยอมรับได้ คือ $.026 < c_j < .317$ ด้วยความเชื่อมั่น 99 %

2. กระบวนการประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้เข้าสอบ กระบวนการประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ คือ การหาค่าประมาณ θ_i $(i = 1, 2,\ldots,N)$ และ b_j, a_j, c_j $(j = 1, 2,\ldots,n)$ ที่ทำให้ฟังก์ชันการแจกแจงภายหลัง $f(\theta, b, a, c/ u)$ มีค่าสูงสุด ปกติจะหาค่า θ_j, b_j, a_j และ c_j ที่ทำให้ $ln\ f(\theta, b, a, c/ u)$ มีค่าสูงสุด ถ้า $ln\ f(\theta, b, a, c/ u)$ เป็นฟังก์ชันที่อนุพันธ์ได้ การหาค่า θ_j, b_j, a_j และ c_j จะหาได้จากการอนุพันธ์ของ $ln\ f(\theta, b, a, c/ u)$ และกำหนดให้อนุพันธ์ของ $ln\ f(\theta, b, a, c/ u)$ มีค่าเท่ากับศูนย์ แล้วหาค่ารากของอนุพันธ์ $ln\ f(\theta, b, a, c/ u)$ ซึ่งแสดงได้ดังสมการต่อไปนี้

$$f(\theta, b, a, c/ u) = L(u/\theta, b, a, c) \cdot f(\theta) \cdot f(b) \cdot f(a) \cdot f(c)/ f(u)$$

$$ln\ f(\theta,b,a,c/u) = ln\ L(u/\theta,b,a,c) \cdot ln\ f(\theta)+ln\ f(b)+ln\ f(a)+ln\ f(c)+\text{constant}$$

$$d\ ln\ f(\theta, b, a, c/ u)\ /d\theta_i = 0$$

$$d\ ln\ f(\theta, b, a, c/ u)\ /db_i = 0$$

$$d\ ln\ f(\theta, b, a, c/ u)\ /da_i = 0$$

$$d\ ln\ f(\theta, b, a, c/ u)\ /dc_i = 0$$

สมการอนุพันธ์ของ $ln\ f(\theta, b, a, c/ u) = 0$ เรียกว่า สมการโมดัล (Model Equation) ค่ารากของสมการโมดัล คือ ค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b, c ที่ทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด อาจทำได้โดยใช้เทคนิค Newton – raphson ซึ่งเป็นการหาค่าประมาณโดยการหาค่า(Iterative) และมีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ของข้อสอบ a_j, b_j และ c_j ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i

$$\theta_i^0 = ln\ (X/\ (n - X_i))$$

เมื่อ ln = Natural Logarithm

X_i = คะแนนสอบของผู้เข้าสอบคนที่ i

n = จำนวนข้อสอบ

ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j, b_j และ c_j

$$a_j^{(0)} = R_j /(1 \quad R_j^2)^{1/2}$$

$$b_j^{(0)} = Z_j / R_j$$

$$c_j^{(0)} = 1 / m_j$$

เมื่อ R_j = Point – biserial Correlation

Z_j = ค่า Z ของการแจกแจงปกติมาตรฐานที่พื้นที่ใต้โค้งปกติมาตรฐานด้านขวามือมีค่าเท่า P_j

$$P_i = \sum_{i=1}^{N} U_{ij} / N$$

เมื่อ N = จำนวนผู้เข้าสอบทั้งหมด

m_j = จำนวนตัวเลือกในข้อสอบข้อที่ j

ขั้นที่ 2 ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$\theta_i^{m+1} = \theta_i^{(m)} - g(\theta_i^m) / h(\theta_i^{(m)})$$

เมื่อ $\theta_i^{(m)}, \theta_i^{(m+1)}$ = ค่าประมาณความสามารถของคนที่ i ครั้งที่ m และ m+1

$g(\theta_i^m)$ = $d\ ln\ f(\theta,b,a,c/u)/d\theta_i$

$h(\theta_i^{(m)})$ = $d^2\ ln\ f(\theta,b,a,c/u)/d\theta_i$

ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ θ_i จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) คือ ค่าประมาณครั้งที่ m และครั้งที่ m+1 มีค่าแตกต่างกันน้อยกว่าค่าคงที่ ที่กำหนดไว้ เช่น .001

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j, b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$a_j^{(m+1)} = a_j^{(m)} - g(a_j^{(m)}) / h(a_j^{(m)})$$
$$b_j^{(m+1)} = b_j^{(m)} - g(b_j^{(m)}) / h(b_j^{(m)})$$
$$c_j^{(m+1)} = c_j^{(m)} - g(c_j^{(m)}) / h(c_j^{(m)})$$

เมื่อ $a_j^m, a_j^{(m+1)}$ = ค่าประมาณอำนาจจำแนกของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$b_j^m, b_j^{(m+1)}$ = ค่าประมาณความยากของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$c_j^m, c_j^{(m+1)}$ = ค่าประมาณการเดาของข้อสอบข้อที่ j ครั้งที่ m และ m+1

$g(a_j^{(m)})$ = $d \ln f(\theta,b,a,c/u)/ da_j$

$h(a_j^{(m)})$ = $d^2 \ln f(\theta,b,a,c/u)/ da_j^2$

$g(b_j^{(m)})$ = $d \ln f(\theta,b,a,c/u)/ db_j$

$h(b_j^{(m)})$ = $d^2 \ln f(\theta,b,a,c/u)/ db_j^2$

$g(c_j^{(m)})$ = $d \ln f(\theta,b,a,c/u)/ dc_j$

$h(c_j^{(m)})$ = $d^2 \ln f(\theta,b,a,c/u)/ dc_j^2$

ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าสู่ค่าคงที่ (Convergence) ค่าอนุพันธ์อันดับที่ 1 $g(x)$ และอนุพันธ์อันดับ 2 $h(x)$ มีส่วนประกอบสองส่วน คือ ส่วนที่ได้จากฟังก์ชันไลค์ลิฮูดและส่วนที่ได้จากการแจกแจงเริ่มแรก ซึ่งส่วนประกอบทั้งสองส่วนนี้แสดงไว้ในตารางที่ 2 – 3

ขั้นที่ 4 ประมาณค่าซ้ำ ขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ θ_i , a_j, b_j และ c_i จะมีค่าคงที่และมีความถูกต้องเพียงพอ หรือทำให้ $f(\theta_i, b, a, c/ u)$ มีค่าสูงที่สุด

ตารางที่ 2 – 4 การอนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 ของ *ln f(θ , b, a, c/ u)*

พารามิเตอร์	อนุพันธ์อันดับที่ 1	อนุพันธ์อันดับที่ 2
θ_i Likelihood	$D\sum_{j=1}^{n} a_j(P_{ij}-c_j)(U_{ij}-P_{ij})$ $/P_{ij}(1-c_j)$	$D^2\sum_{j=1}^{n} a^2,(P_{ij}-C_j)(U_{ij}C_j-P_{ij}^2)$ $Q_{ij}/P_{ij}^2(1-C_j)^2$
Prior	$-\theta_i$	-1
a_j Likelihood	$D\sum_{j=1}^{n}(\theta_i-b)(P_{ij}-C_{ij})$ $(U_{ij}-P_{ij})/P_{ij}(1-C_j)$	$D^2\sum_{j=1}^{n}(\theta_i-b_j)^2(P_{ij}-C_{ij})$ $(U_{ij}C_j-P_{ij}^2)Q_{ij}/P_{ij}^2(1-/c_j)^2$
Prior	$+(v_j-1)/a_j-a_j/w_j$	$-(v_j-1)/a_j^2-1/w_j$
b_j Likelihood	$-D\sum_{j=1}^{n} a_j(P_{ij}-C_j)(U_{ij}-P_{ij})$ $/P_{ij}(1-c_j)$	$D^2\sum_{j=1}^{n} a_j(P_{ij}-C_j)(U_{ij}C_j-P_{ij})$ $Q_{ij}/P_{ij}^2(1-c_j)^2$
Prior	–	–
c_j Likelihood	$\sum_{i=1}^{N}(U_{ij}-P_{ij})/P_{ij}(1-C_j)$	$\sum_{i=1}^{N}[U_{ij}2P_{ij}-1)-P_{ij}^2/P_{ij}^2(1-C_j)^2$
Prior	$+s_j/c_j-t_j/(1-c_j)$ $P_{ij}=c_j+(1-c_j)/(1+\exp(-Da_j$ $(\theta_i-b_j)))$	$-s_j/c_j^2+t_j/(1-c_j)^2$

การทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยโปรแกรม WinBUGS

จากการศึกษาของ Saengla Chaimongkol et al. (2007) ได้ประยุกต์วิธีประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดลพหุระดับแบบ 3 ระดับ ได้แก่ ระดับที่ 1 (ระดับข้อสอบ) ระดับที่ 2 (ระดับนักเรียน) และระดับที่ 3 (ระดับโรงเรียน)

ให้ Y_{ijk} คือ ผลการตอบข้อสอบถูกในข้อที่ *k* ของคนที่ *j* โรงเรียนที่ *i*

S_i คือ โรงเรียนที่ *i*

เมื่อ Y_{ijk} มีค่าเป็น 1 และ 0 ดังนั้น จึงมีการแจกแจงแบบเบอร์นุยลี (Bernoulli)

Y_{ijk} ~ Bernoulli (P_{ijk})

โมเดลการตอบข้อสอบข้อที่ *k* ของคนที่ *j* โรงเรียนที่ *i* ที่ข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยให้ *G* เป็นตัวแปรกลุ่ม เป็นดังนี้

$$logit(P_{ijk}) = u3_i + u2_{ij} + \alpha 0\ G_{ij} - \beta_k - \gamma 0_k G_{ij} + \delta 0_k S_i + \delta 1_k S_i G_{ij} + u4_{ik} G_{ij}$$

u_{3i} คือ อิทธิพลสุ่มของโรงเรียน i ที่มีการแจกแจงเป็นโค้งปกติ มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเท่ากันทุกโรง $(\sigma 3)^2$

u_{2i} คือ ความสามารถของนักเรียนในโรงเรียน เป็นอิทธิพลสุ่มและมีการแจกแจงเป็นโค้งปกติ ค่าเฉลี่ยไม่เป็น 0 ความแปรปรวนของกลุ่มแต่ละกลุ่มเป็น σ_G^2

$\alpha 0$ คือ อิทธิพลคงที่ที่เป็นผลการเปรียบเทียบระหว่างกลุ่มเปรียบเทียบกับกลุ่มอ้างอิงเป็นผลต่างของค่าเฉลี่ยระหว่างกลุ่มเปรียบเทียบกับกลุ่มอ้างอิง

G_{ij} คือ ตัวแปรกลุ่มโดยแบ่งออกเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ G_{ij} มีค่าเป็น 0 ถ้านักเรียนคนที่ *j* ในโรงเรียนที่ *i* อยู่ในกลุ่มอ้างอิง และมีค่าเป็น 1 ถ้านักเรียนคนที่ *j* ในโรงเรียนที่ *i* ของกลุ่มเปรียบเทียบ

β_k คือ อิทธิพลคงที่แทนค่าความยากของข้อสอบข้อที่ *k* ของกลุ่มอ้างอิง

$\gamma 0_k$ คือ ค่าเฉลี่ยรวมของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ข้อที่ *k* ระหว่างโรงเรียน

$\delta 0_k$ คือ อิทธิพลคงที่เป็นอิทธิพลหลักของของตัวแปรระดับโรงเรียนของข้อที่ *k*

$\delta 1_k$ คือ อิทธิพลคงที่เป็นค่าแสดงผลของปฏิสัมพันธ์ระหว่างระดับโรงเรียนกับกลุ่มผู้สอบของข้อที่ *k* กล่าวในอีกความหมายหนึ่ง $\delta 1_k$ เป็นค่าที่บอกถึงขนาดการทำหน้าที่ต่างกันของข้อสอบจะแตกต่างกันขึ้นอยู่กับลักษณะของโรงเรียน

$u4_{ik}$ คือ อิทธิพลสุ่มที่เพิ่มขึ้นของค่าการทำหน้าที่ต่างกันของข้อที่ *k* ในโรงเรียน *i* มีการแจกแจงเป็นโค้งปกติมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเฉพาะข้อเป็น $(\sigma 4_i)^2$ และมีข้อตกลงต่ออีกว่าอิทธิพลสุ่ม u_{3i} , u_{2i} เป็นอิสระต่อกัน

ค่า $\sigma 4_k$ จะนำไปสู่ดัชนีต่าง ๆ ในการวัดการทำหน้าที่ต่างกันของข้อสอบระหว่างโรงเรียน ถ้าค่า $\sigma 4_k$ ชี้ให้เห็นว่า หลังจากควบคุมให้ความสามารถของโรงเรียนและของนักเรียนแล้วค่าการทำหน้าที่ต่างกันของข้อสอบยังมีค่าแปรเปลี่ยนระหว่างโรงเรียนมาก ในทางตรงข้าม ถ้าค่า $\sigma 4_k$ เป็น 0 หรือใกล้ ๆ 0 แสดงว่าการเกิด DIF มีการแปรเปลี่ยนระหว่างโรงเรียนน้อย ในสมการที่กล่าวข้างต้น ยังไม่สามารถหาคำตอบที่เหมาะที่สุดเป็นค่าเดียวได้ (Identified) เนื่องจากสามารถเพิ่มค่าคงที่เข้ากับ u_{3i} และทุก ๆ β_k แต่ค่าโลจิตของโมเดลไม่มีค่าเปลี่ยนแปลง นอกจากนั้น $\delta 0_k$ ยังมีปัญหาทำให้เกิดปัญหาการหาคำตอบที่เหมาะที่สุดเป็นค่าเดียวได้เช่นกัน เนื่องจากเราสามารถลบค่าคงที่ *c* ออกจากค่า $\delta 0_k$ ทุก ๆ *i* และเพิ่มเทอม cS_i เข้ากับ u_{3i} โดยไม่มีผลต่อค่าโลจิต

เพื่อให้โมเดลข้างต้นสามารถหาคำตอบได้เราจะใช้แนวคิดของ Bafurmi et al. (2005) โดยการแทนที่ค่าพารามิเตอร์ด้วยการปรับแก้บางค่าได้แก่ ปรับแก้ $\delta 0_k$ โดยลบด้วย $\bar{\delta} 0$ และบวก $\bar{\delta} 0 S_i$ เข้ากับเทอม u_{3i} ทุก ๆ ค่า ถ้าเราเปลี่ยนค่าศูนย์กลางของ S_i ให้มีค่าเฉลี่ยเป็น 0 การบวก $\bar{\delta} 0 S_i$ จะไม่มีผลต่อค่าเฉลี่ยของ u_{3i} เทอมความคลาดเคลื่อนจะปรับแก้เป็นดังนี้

$$u3_i^{adj} = u3_i - \bar{u}3 + \bar{\delta} 0 S_i$$
$$u2_{ij}^{adj} = u2_{ij} - \bar{\beta} + \bar{u}3$$
$$\alpha 0^{adj} = \alpha 0 \quad \bar{\gamma} 0$$
$$\gamma 0_k^{adj} = \gamma 0_k - \bar{\gamma} 0$$
$$\delta 0_k^{adj} = \delta 0_k - \bar{\delta} 0$$

เมื่อแทนค่าพารามิเตอร์เดิมด้วยค่าที่ปรับแก้ จะได้โมเดล ดังนี้

$$logit(P_{ijk}) = u3_i^{adj} + u2_{ij}^{adj} + \alpha 0^{adj} G_{ij} - \beta_k^{adj} - \gamma 0_k^{adj} G_{ij} + \delta 0_k^{adj} S_i + \delta 1_k S_i G_{ij} + u4_{ik} G_{ij}$$

โมเดลนี้ ค่า $\gamma 0_k^{adj}$ ที่เป็นสัมประสิทธิ์ของตัวแปรกลุ่ม G_{ij} ถือเป็นค่า DIF ของข้อสอบข้อที่ k

β_k^{adj} คือ ค่าความยากของข้อสอบ

$\alpha 0^{adj}$ คือ ค่าผลต่างของค่าเฉลี่ยระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ

$u3_i^{adj}$ คือ ความสามารถของโรงเรียน

$u2_{ij}^{adj}$ คือ ความสามารถของนักเรียน

$u4_{ik}$ คือ ผลอิทธิพลสุ่มของการเกิด DIF

$\delta 0_k^{adj}$ คือ เป็นอิทธิพลหลักของโรงเรียนที่ปรับค่าแล้ว

ส่วนเทอมอื่น ๆ ที่เหลือมีความหมายเช่นเดียวกับโมเดลที่ยังไม่ได้ปรับแก้

การประมาณค่าพารามิเตอร์ของโมเดลโดยใช้โปรแกรม WinBUGS

โปรแกรม WinBUGS สามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้โดยวิธีเบส์เซียน ที่ใช้การสุ่มตัวอย่างในการจำลองข้อมูลแบบกิบส์ (Gibbs Sampling) พารามิเตอร์ที่ต้องการวิเคราะห์ได้แก่ ขนาดของ $\delta 0_k$, $\bar{\delta} 0$ และ $\delta 1_k$ ถ้าค่าของ $\delta 1_k$ มีค่ามาก เทอม $\delta 0_k S_i$ จะบ่งบอกความแปรเปลี่ยนของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างโรงเรียน การเพิ่มเทอมนี้ในโมเดลจะทำให้ลดค่าของ $\sigma 4_k$

วิธีเบส์เซียนจะถือว่าค่าพารามิเตอร์ที่ไม่ทราบค่าทั้งหมดเป็นค่าเชิงสุ่มที่มีการแจกแจงเริ่มต้น (Prior Distribution) ที่เหมาะสม การประมาณค่าอาศัยการแจกแจงภายหลังร่วม (Joint Posterior Distribution) $P(\theta|y)$ เมื่อ θ เป็นเวคเตอร์ค่าพารามิเตอร์ที่ไม่ทราบค่า เช่น $(\theta(\{\alpha 0\},\{\beta_k\},\{\gamma 0_k\},\{\delta 0_k\},\{\delta 1_k\},\mu,\{\sigma 2_G\},\sigma 3,\{\sigma 4_k\}))$ และ y เป็นข้อมูลจากกลุ่มตัวอย่าง

การแจกแจงภายหลังของ θ โดยทฤษฎีของเบส์ จะได้ค่าเป็น

$$P(\theta|y)=\frac{P(y|\theta)P(\theta)}{\int P(\theta)P(y|\theta)d\theta} \propto P(y|\theta)P(\theta)$$

เมื่อ $P(\theta|y)$ คือ ไลค์ลิฮูด และ $P(\theta)$ เป็นการแจกแจงเริ่มต้น การแจกแจงภายหลังเป็นสัดส่วนของไลค์ลิฮูดคูณกับการแจกแจงเริ่มต้น

เนื่องจากการตอบข้อสอบ โดยทราบค่าความสามารถของโรงเรียนและของนักเรียนมีความเป็นอิสระกัน ค่าไลค์ลิฮูดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นดังนี้

$$P(y|\theta)=\int g(u4;0,\sigma 4_k)\int g(u3;0,\sigma 3)$$

$$\int g(u2;\mu,\sigma 2_G)\prod_{i,j,k}\int y_{ijk}|u2_{ij},u3_i u4_{ik})\ du2_{ij}\,du3_i\,du4_{ik},$$

เมื่อ $f(y_{ijk}|u2_{ij},u3_i,u4_{ik})=\left(\frac{1}{1+e^{-n_{ijk}}}\right)^{y_{ijk}}\left(\frac{e^{-n_{ijk}}}{1+e^{n_{ijk}}}\right)^{1-y_{ijk}}$

นอกจากนี้ วิธีเบส์เซียน มีความแตกต่างจากวิธีดั้งเดิมในการอนุมานรูปแบบความน่าจะเป็นของโมเดล สำหรับตัวแปรสังเกตและพารามิเตอร์ที่ไม่ทราบค่าเบส์เซียนจะอนุมานโดยการตรวจสอบเงื่อนไขของพารามิเตอร์ที่สอดคล้องกับข้อมูลที่สังเกตได้จึงสามารถใช้งานได้ง่าย วิธีเบส์เซียนมีความยืดหยุ่นสูง สามารถแก้ปัญหาทั้งง่ายและซับซ้อนได้ดี มีการกำหนดการแจกแจงเริ่มต้นของค่าพารามิเตอร์ (Prior Distribution) ที่ใช้ในการกำหนดช่วงของค่าพารามิเตอร์ที่ต้องการประมาณค่า เป็นประโยชน์อย่างยิ่งสำหรับการวิเคราะห์ข้อมูล นอกจากนั้นโมเดลที่มีความซับซ้อน ยิ่งในปัจจุบันการพัฒนาเทคนิคการสุ่มตัวอย่างในการจำลองข้อมูล Markov Chain Monte Carlo (MCMC) มีโปรแกรมสนับสนุนการประมาณค่ามากมาย เช่น โปรแกรม WinBUGS ซึ่งเป็นซอฟแวร์ที่ใช้งานได้ง่าย

ตอนที่ 6 การทดสอบวัดสัมฤทธิ์ระดับชาติของนักเรียน

ความหมายของการทดสอบวัดผลสัมฤทธิ์ระดับชาติของนักเรียน (ศศิธร สุริยา, 2551)

นักวิชาการได้ให้ความหมายของการวัดผลสัมฤทธิ์ระดับชาติไว้หลายท่านดังนี้

สำนักทดสอบการศึกษา (2547, หน้า 25) ได้ให้ความหมาย การวัดผลสัมฤทธิ์ระดับชาติ คือ การสอบประเมินคุณภาพการศึกษาระดับชาติขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ตรวจสอบกำกับดูแล และพัฒนาคุณภาพการศึกษาของโรงเรียนของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) กระทรวงศึกษาธิการ

นิพล พลกลาง (2549, หน้า 45 – 46) ได้ให้ความหมาย การวัดผลสัมฤทธิ์ระดับชาติ คือ การสอบประจำปีที่ทางกระทรวงศึกษาธิการให้เด็กไทยทุกคนต้องสอบในทุกปี โดยนักเรียนระดับชั้นประถมศึกษาปีที่ 3 ประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยม ศึกษาปีที่ 6 จะต้องสอบจากข้อสอบมาตรฐานจากส่วนกลาง โดยมีกรรมการคุมการสอบจากโรงเรียนอื่น

จากความหมายต่าง ๆ สรุปได้ว่า การวัดผลสัมฤทธิ์ระดับชาติ คือ คะแนนที่ได้จากการสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชั้นประถมศึกษาปีที่ 3 ประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 เพื่อนำไปประเมินคุณภาพการศึกษาตามมาตรฐานการศึกษา ประกอบด้วย วิชา ภาษาไทย สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ และภาษาอังกฤษ โดยกรมวิชาการ กระทรวงศึกษาธิการ

ความเป็นมาของการจัดสอบวัดผลสัมฤทธิ์ระดับชาติ

กระทรวงศึกษาธิการมีนโยบายที่จะดำเนินการปฏิรูปการศึกษาให้สอดคล้องกับเจตนารมณ์ แห่งรัฐธรรมนูญและพระราชบัญญัติการศึกษาแห่งชาติ โดยมุ่งที่จะพัฒนาประเทศไทยให้เป็นสังคมแห่งการเรียนรู้อันเป็นเงื่อนไขสำคัญที่จะนำไปสู่ระบบเศรษฐกิจบนฐานความรู้ ให้ชาวไทยทุกคนมีโอกาสเท่าเทียมกันที่จะได้รับความรู้และการฝึกอบรมตลอดชีวิตมีปัญญาเป็นทุนไว้สร้างงาน สร้างรายได้ และสามารถนำประเทศให้ก้าวพ้นวิกฤติทางเศรษฐกิจและสังคม พระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 การปฏิรูปการศึกษาและกระทรวงศึกษาธิการ มีภารกิจด้านการวัดและประเมินผลที่ต้องปรับเปลี่ยนไปจากที่เคยดำเนินการทั้งด้านขอบเขต สาระและวิธีการดำเนินการซึ่งจะนำไปสู่การพัฒนาโครงสร้างขององค์กรและโครงการเพื่อรองรับและสนับสนุนการปฏิบัติภารกิจของการะทรวงศึกษาธิการด้านการวัดและประเมินผล ซึ่งมีลักษณะสำคัญดังนี้

1. เป็นการประเมินมาตรฐานการศึกษาของชาติ เพื่อควบคุมคุณภาพการศึกษาและเพื่อรักษาความมีเอกภาพด้านนโยบาย

2. เป็นการประเมินคุณภาพผู้เรียนเพื่อไม่ให้เกิดการพัฒนาระดับปัจเจกบุคคลในองค์รวมอย่างต่อเนื่อง และสอดคล้องกับสภาพจริง

3. เป็นการประเมินเพื่อสนับสนุนหลักการจัดการศึกษาเพื่อสร้างชาติ สร้างคน และสร้างงานตามนโยบายของรัฐบาล

การทดสอบวัดผลสัมฤทธิ์ระดับชาติในต่างประเทศ

ในขณะที่การปฏิรูปการศึกษากำลังเป็นหนึ่งในกระแสหลักของโลกาภิวัฒน์ การประเมินมาตรฐานการศึกษาระดับประเทศโดยใช้แบบทดสอบมาตรฐานได้เกิดขึ้นในประเทศส่วนใหญ่ของโลก ยกเว้นบางประเทศ เช่น Iceland, Sweden เป็นต้น ที่มีกระแสต่อต้านการตรวจสอบควบคุมการจัดการศึกษาจากส่วนกลาง (Stake, 1955 อ้างถึงใน นิพล พลกลาง, 2549, หน้า 15) จากการศึกษาค้นคว้าเอกสารจาก Websites ของหน่วยงานที่รับผิดชอบการประเมินของประเทศต่าง ๆ เช่น สหรัฐอเมริกา (NAGB, 2000 อ้างถึงใน นิพล พลกลาง, 2549, หน้า 15) สหราชอาณาจักร (QCA, 2001) ออสเตรเลีย (New South Wale Board of Student, 2001; Victoria Board of Student, 2001 อ้างถึงใน นิพล พลกลาง, 2549, หน้า 15) และสิงคโปร์ (Singapore Misnistry of Education, 2001 อ้างถึงใน นิพล พลกลาง, 2549, หน้า 15) พบว่า มีการประเมินระดับชาติ 2 รูปแบบใหญ่ คือ

1. การประเมินระดับชาติที่ไม่มีผลต่อการตัดสินผลการเรียน เป้าหมายของการประเมินแบบนี้ คือ การได้ข้อมูลเกี่ยวกับคุณภาพการศึกษาของชาติในลักษณะภาพรวมผล การประเมินสามารถแสดงให้เห็นคุณภาพการศึกษา ณ เวลาที่ประเมินและแสดงแนวโน้มในระยะยาว โดยการวิเคราะห์เปรียบเทียบตัวอย่างการประเมินแบบนี้ ได้แก่ National Assessment of Educational Progress (NAEP) ของสหรัฐอเมริกา และ Qualifications and Curriculum Authority's (QCE) National Test ของสหราชอาณาจักร แต่ QCE และ NAEP ก็มีความแตกต่างที่สำคัญ คือ NAEP เป็นการสอบเฉพาะผู้ที่เป็นกลุ่มตัวอย่าง ส่วน QCE เป็นการสอบสำหรับนักเรียนทุกคนที่มีอายุครบ 7 ปี 11 ปี และ 14 ปี ผลการสอบของ QCE จึงสามารถใช้แสดงผลสัมฤทธิ์ระดับรายบุคคลได้ ซึ่งเป็นข้อมูลที่มีประโยชน์อย่างยิ่งต่อผู้เรียน ครูผู้สอน และผู้ปกครองที่จะร่วมกันพัฒนาคุณภาพผู้เรียนแต่ละคน

2. การประเมินผลระดับชาติ เพื่อตัดสินผลการเรียนของผู้เรียนระดับตัวประโยคต่าง ๆ เช่น การสอบ O – level และ A – level ของสหราชอาณาจักร การสอบ Higher School

Certificate ในรัฐ New South Wales ประเทศ Australia (GNE) N – level, GNEO – level และ GNEA – level ของประเทศสิงคโปร์ การประเมินรูปแบบนี้ แม้จะเกิดในประเทศที่จัดการศึกษาแบบกระจายอำนาจแต่รัฐยังสงวนไว้ซึ่งอำนาจการประเมินผลและการตัดสิน

การทดสอบวัดผลสัมฤทธิ์ระดับชาติในประเทศไทย

การทดสอบวัดผลสัมฤทธิ์ระดับชาติ หรือ National Test นับตั้งแต่กระทรวงศึกษาธิการได้กระจายอำนาจการวัดและประเมินผลการศึกษาสู่ระดับสถานศึกษา การประเมินผลสัมฤทธิ์ผู้เรียนทั้งในระดับการเรียนการสอน การตัดสินได้ – ตก และการอนุมัติการสำเร็จการศึกษาเป็นอำนาจของสถานศึกษา ถึงแม้ว่าการจัดการเรียนการสอนของสถานศึกษาทั่วประเทศจะใช้หลักสูตรเดียวกันของกระทรวงศึกษาธิการ แต่ความหลากหลายในทางปฏิบัติในเรื่องรูปแบบการสอน การใช้สื่อและอุปกรณ์การเรียนการสอน ตลอดจนการวัดและประเมินผลได้เกิดขึ้นอย่างหลีกเลี่ยงไม่ได้ ทั้งในสถานศึกษาต่าง ๆ ภายในภูมิภาคเดียวกันและต่างภูมิภาค เพื่อเป็นการควบคุมและรักษาคุณภาพการศึกษาของสถานศึกษาต่าง ๆ ทั่วประเทศให้มีมาตรฐานใกล้เคียงกัน กระทรวงศึกษาธิการจึงได้จัดให้มีการประเมินคุณภาพการจัดการศึกษาระดับชาติ ผลการประเมินที่ได้นอกจากจะเป็นตัวบ่งชี้คุณภาพการศึกษาของประเทศในภาพรวมแล้ว ยังใช้เป็นข้อมูลประกอบการตัดสินใจในการกำหนดนโยบายระดับชาติ และใช้วางแผนในการพัฒนาคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา และระดับโรงเรียนการพัฒนาการด้านการวัดและประเมินผลในวงการศึกษาไทย และข้อกำหนดพระราชบัญญัติการศึกษาแห่งชาติเกี่ยวกับการตัดสินผลการเรียน ซึ่งเป็นอำนาจหน้าที่ของสถานศึกษา ดังนั้น การทดสอบวัดผลสัมฤทธิ์ระดับชาติในรูปแบบการสอบ O – level และ A – level ในประเทศอังกฤษหรือการจัดสอบ Exit Examination ในบางมลรัฐในประเทศสหรัฐอเมริกา คงเป็นทางเลือกที่เป็นไปไม่ได้สำหรับรูปแบบที่กระทรวงศึกษาธิการใช้อยู่ในปัจจุบันก็ค่อนข้างคล้ายคลึงกับ NAEP ของสหรัฐอเมริกา

แนวทางการประเมินคุณภาพการศึกษา

ในบทบัญญัติพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 มาตรา 4 ได้ให้นิยามศัพท์ที่เกี่ยวข้องกับคำว่า มาตรฐานการศึกษา หมายถึง ข้อกำหนดที่เกี่ยวกับคุณลักษณะ คุณภาพที่พึงประสงค์และมาตรฐานที่ต้องการให้เกิดขึ้นภายในสถานศึกษาทุกแห่งและเพื่อให้หลักในการเทียบเคียงสำหรับส่งเสริม กำกับดูแล การตรวจสอบ การประเมินผล และการประกันคุณภาพการศึกษาตามนัยแห่งความหมายของมาตรฐานการศึกษาดังกล่าว สามารถแยกเป็นคำสำคัญ 3 คำ ได้แก่

1. คุณลักษณะ หมายถึง สิ่งที่เป็นลักษณะสำคัญของการศึกษา

2. คุณภาพ หมายถึง คุณภาพของคุณลักษณะดังกล่าว เช่น คุณภาพสูง คุณภาพต่ำ ในนิยามนี้ คำว่า คุณภาพที่พึงประสงค์ หมายถึง ที่พึงประสงค์ของสังคม ซึ่งผู้จัดต้องกำหนดขึ้นมาว่าคืออย่างไร

3. มาตรฐาน หมายถึง ความมีบรรทัดฐานที่ยอมรับกันให้เป็นมาตรวัด การกำหนดมาตรฐานกำหนดขึ้นโดยผู้รับผิดชอบ (วิชัย ตันศิริ, 2543, หน้า 48 อ้างถึงใน นิพล พลกลาง, 2549, หน้า 21) ซึ่งจากนิยามในพระราชบัญญัติ สรุปได้ว่า การสร้างความเป็นมาตรฐานการศึกษาของชาตินั้น จะมีปัจจัยสำคัญที่เป็นองค์ประกอบคุณลักษณะทางการศึกษาที่ก่อให้เกิดคุณภาพที่พึงประสงค์ภายใต้มาตรฐานที่ยอมรับร่วมกัน ซึ่งเรียกรวมกันว่าเป็นการสร้างความเป็นคุณภาพการศึกษา คำว่า คุณภาพ (Quality) เป็นคำที่ใช้กันมาก โดยเฉพาะในวงการธุรกิจ ซึ่งมีผู้ให้ความหมายไว้ต่างกัน และในสภาพปัจจุบันจะหมายถึงการทำให้ลูกค้าพึงพอใจด้วยการทำให้ความต้องการและความหวังของลูกค้าได้รับการตอบสนอง เช่น คุณภาพในการศึกษาประทับใจหรือมั่นใจในคุณภาพของผลผลิต คือ นักเรียนมีคุณภาพตามมาตรฐานที่กำหนดการประเมินคุณภาพการศึกษาเป็นกลไกหนึ่งที่มีความสำคัญซึ่งส่งผลต่อการจัดระบบการประกันคุณภาพการศึกษาของสถานศึกษา การประเมินผลที่ดีสามารถบ่งบอกถึงผลการดำเนินงานได้อย่างถูกต้อง บ่งบอกถึงระดับความมั่นใจและพึงพอใจของสถานศึกษา และชุมชนและสาธารณชนทั่วไป การจัดการศึกษาตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 สถานศึกษาต้องนำหลักสูตรไปใช้พัฒนาผู้เรียนให้เกิดกรเรียนรู้อย่างมีคุณภาพตามมาตรฐานหลักสูตรและมีคุณลักษณะตามจุดหมายหลักสูตรและให้สถานศึกษาจัดให้ผู้เรียนทุกคนในทุกระดับช่วงชั้นที่หลักสูตรการศึกษาขั้นพื้นฐานกำหนด โดยสถานศึกษาประเมินคุณภาพการศึกษาในลักษณะต่าง ๆ ดังนี้

1. ประเมินผลสัมฤทธิ์ในวิชาแกนหลักและคุณลักษณะที่สำคัญด้วยเครื่องมือมาตรฐานดังนี้

1.1 ประเมินผลการเรียนรู้ ผู้เรียนตามกลุ่มสาระการเรียนรู้ทั้ง 8 กลุ่ม

1.2 ประเมินการอ่าน คิด วิเคราะห์ เขียน

1.3 ประเมินคุณลักษณะอันพึงประสงค์

1.4 ประเมินผลการเข้าร่วมกิจกรรมพัฒนาผู้เรียน

2. การประเมินคุณภาพการศึกษาตามมาตรฐานการศึกษาระดับสถานศึกษาทุกมาตรฐาน มีขั้นตอนการประเมิน ดังนี้

2.1 กำหนดวัตถุประสงค์ของการประเมินคุณภาพการศึกษาตามมาตรฐานการศึกษาระดับสถานศึกษาและตัวบ่งชี้ที่สถานศึกษาต้องการประเมินผล

2.2 กำหนดกรอบการประเมิน ซึ่งครอบคลุมตัวบ่งชี้ วิธีการประเมิน เครื่องมือที่ใช้ในการประเมินผลและแหล่งข้อมูล

2.3 กำหนดเกณฑ์การตัดสินผลการดำเนินงานของสถานศึกษา โดยเปรียบเทียบกับเกณฑ์มาตรฐานที่กำหนด

2.4 ดำเนินการประเมินคุณภาพการศึกษา

2.5 วิเคราะห์ข้อมูลที่ได้มาจากการจัดเก็บจากแต่ละแหล่งข้อมูล

2.6 สรุปและรายงานผลการประเมินคุณภาพการศึกษา

การประเมินคุณภาพการศึกษาระดับชาติ

การประเมินคุณภาพการศึกษาระดับชาติ มีอยู่ 2 ลักษณะ คือ

1. การทดสอบวัดผลสัมฤทธิ์ทางการเรียน (General Achievement Test) หรือ GAT เป็นการประเมินที่มีจุดมุ่งหมายเพื่อประเมินผลสัมฤทธิ์ทางการเรียนของผู้เรียนตามระดับช่วงชั้นที่หลักสูตรกำหนดได้แก่ชั้นประถมศึกษาปีที่ 3 ชั้นประถมศึกษาปีที่ 6 ช่วงชั้นมัธยมศึกษาปีที่ 3 ชั้นมัธยมศึกษาปีที่ 6 โดยใช้แบบทดสอบที่พัฒนาเป็นแบบสอบมาตรฐาน (Standardized Test) เป็นเครื่องมือสำคัญในการประเมิน ลักษณะของข้อสอบเป็นแบบทดสอบปรนัยชนิดเลือกตอบ โดยมีสาระการเรียนรู้การประเมิน ดังนี้ ประถมศึกษาปีที่ 3 ประเมินใน 3 วิชา คือ ภาษาไทย คณิตศาสตร์ ประถมศึกษาปีที่ 6 ประเมินใน 3 วิชา คือ ภาษาไทย คณิตศาสตร์ และภาษาอังกฤษ มัธยมศึกษาปีที่ 3 ประเมิน 5 รายวิชา ได้แก่ ภาษาไทย คณิตศาสตร์ ภาษาอังกฤษ วิทยาศาสตร์ และสังคมศึกษา ศาสนาและวัฒนธรรม และชั้นมัธยมศึกษาปีที่ 6 ประเมินสาระการเรียนรู้ทั้ง 8 กลุ่มสาระตามหลักสูตรการศึกษา ขั้นพื้นฐานพุทธศักราช 2544 (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการศึกษาขั้นพื้นฐาน, 2549, หน้า 1 – 2)

2. การทดสอบวัดผลความถนัดทางการเรียน (Scholastic Aptitude Test) หรือ SAT เป็นการประเมินความถนัดทางการเรียนของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ทุกคนในสถานศึกษาที่อยู่ในระบบทุกสังกัดทั่วประเทศเป็นการประเมินที่วัดความสามารถของผู้เรียนที่ไม่ใช่การวัดความรู้ด้านเนื้อหาสาระตามหลักสูตรเพียงอย่างเดียวแต่เป็นการวัดสิ่งที่ตกตะกอนอยู่ในตัวผู้เรียนอันเป็นผลที่เกิดจากการสะสมระยะยาวของกระบวนการเรียนรู้ที่เกิดขึ้นตามหลักสูตรทั้งในและนอกห้องเรียน และเป็นความรู้ความสามารถที่ผ่านกระบวนการกลั่นกรองสังเคราะห์มาเป็นเวลายาวนานหลาย ๆ ปี ซึ่งจะเป็นตัวบ่งชี้ให้เห็นถึงความสามารถของคนที่ได้รับการพัฒนาแล้วและเป็นตัวพยากรณ์ที่ศึกษาความสำเร็จทางการศึกษาในอนาคตเครื่องมือที่ใช้

ในการประเมินเป็นแบบทดสอบความถนัดทางการเรียนใน 3 องค์ประกอบ ความสามารถทางภาษา (Verbal Ability) เป็นข้อสอบที่วัดความรู้ ความเข้าใจ และเหตุผลทางภาษาของผู้เรียน โดยครอบคลุมพื้นฐาน ความรู้ความเข้าใจทางภาษา ความสามารถในการวิเคราะห์ เปรียบเทียบและประเมินความสัมพันธ์รูปแบบต่าง ๆ ระหว่างคำศัพท์ สำนวน ข้อมูลและแนวความคิดที่สื่อในรูปข้อความ หรือบทความ เหตุผลเชิงภาษา (Verbal Reasoning) และการอ่านอย่างมีวิจารณญาณ (Critical Reading) ความสามารถทางการคิดคำนวณ (Numerical Ability) เป็นข้อสอบที่วัดความเข้าใจในความคิดรวบยอดและหลักการทางคณิตศาสตร์ระดับเบื้องต้นลักษณะการคำนวณระดับ พื้นฐานความสามารถด้านเหตุผลเชิงปริมาณ (Quantitative Reading) การวิเคราะห์เปรียบเทียบและประเมินข้อมูลเชิงปริมาณในรูปแบบต่าง ๆ และความสามารถในการแก้ปัญหาในรูปของจำนวนหรือปริมาณความสามารถเชิงวิเคราะห์ เป็นข้อสอบที่วัดความสามารถของผู้เรียนด้านต่าง ๆ เช่น การสังเกตเห็นและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลสารสนเทศต่าง ๆ การพิจารณาประเมินความสอดคล้องระหว่างข้อความที่มีความสัมพันธ์เกี่ยวข้องกัน การสรุปความจากข้อมูลหรือสาระที่ซับซ้อน การใช้วิธีหรือขั้นตอนในการขจัดประเด็นหรือทางเลือกที่ผิดเพื่อนำไปสู่ผลการสรุปอย่างถูกต้อง การสรุปสร้างหลักเกณฑ์จากความสัมพันธ์ที่วิเคราะห์จากข้อมูลที่เป็นนามธรรมสัญลักษณ์ หรือแผนภาพต่าง ๆ การวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มหรือการจัดการประเภทที่มีความเลื่อมล้ำในลักษณะต่าง ๆ หรือที่เป็นอิสระไม่คาบเกี่ยวกัน เป็นต้น (นิพล พลกลาง, 2549, หน้า 19 – 20)

ความสำคัญของการทดสอบวัดผลสัมฤทธิ์ระดับชาติ

การดำเนินการประเมินคุณภาพการศึกษาขั้นพื้นฐาน ทำหน้าที่เป็นกลไกการควบคุมคุณภาพการศึกษาคุณภาพของผู้เรียนเป็นเครื่องมือกระตุ้น สร้างแรงจูงใจในผลสัมฤทธิ์ ผลักดันคุณภาพของงานให้เกิดขึ้นทั้งกับครูผู้สอนและตัวผู้เรียนอีกทางหนึ่งด้วย ดังนั้นผลที่เกิดขึ้นกับผู้เรียนจากการจัดการศึกษาที่มีคุณภาพควรสะท้อนให้เห็นได้จากคะแนนการทดสอบด้วย แบบทดสอบ หรือ เครื่องมือวัดประเภทต่าง ๆ ที่สูงขึ้น หรือมีพัฒนาการที่ดีขึ้น ดังนั้นการประเมินคุณภาพการศึกษาระดับชาติ หรือการทดสอบวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนจึงเป็นเครื่องมือสำคัญยิ่งในสร้างความมั่นใจเกี่ยวกับคุณภาพการศึกษา ทั้งนี้เพราะการประเมินระดับชาติจะทำหน้าที่เป็นมาตรฐานกลางที่ใช้เทียบเคียงผลที่เกิดขึ้น อีกทั้งจะทำให้ได้ข้อมูลซึ่งเป็นตัวบ่งชี้คุณภาพการศึกษาของชาติในภาพรวม ทำให้ได้ข้อมูลประกอบการตัดสินใจเชิงนโยบาย และการวางแผนพัฒนาคุณภาพการศึกษาของกระทรวงศึกษาธิการโดยที่ข้อมูล

ผลการประเมินจะถูกนำเสนอหลายระดับ ได้แก่ ระดับผู้เรียนรายบุคคลรายสถานศึกษา เขตพื้นที่การศึกษา ระดับประเทศ และจำแนกสังกัด กรณีมีสถานศึกษาสังกัดอื่นอาทิ โรงเรียนสาธิต เทศบาล เอกชน หรือ กรุงเทพมหานครฯ เข้าร่วมการประเมิน ดังนั้นผู้เกี่ยวข้อง หรือผู้มีส่วนได้ส่วนเสีย แม้กระทั่งผู้สนใจศึกษาสามารถใช้ประโยชน์จากข้อมูลที่รายงานแต่ละระดับได้ข้อมูลระดับผู้เรียนรายบุคคล ซึ่งผู้ใช้อาจ ได้แก่ ครูผู้สอน ครูแนะแนว โรงเรียนผู้ปกครอง ผู้บริหารสถานศึกษา และตัวนักเรียนเอง

ข้อมูลระดับนี้สามารถนำไปใช้เพื่อแสดงระดับผลสัมฤทธิ์ของผู้เรียนเป็นข้อมูลเพื่อการพัฒนาผู้เรียนรายบุคคลอย่างต่อเนื่อง เป็นข้อมูลพื้นฐานเพื่อประเมินนักเรียน/ โรงเรียนเพื่อการคัดเลือกให้รางวัลศึกษาต่อหรือรองรับคุณภาพใช้เป็นข้อมูลเพื่อการสื่อสาร ให้ทราบถึงคุณภาพนักเรียน/ศักยภาพของโรงเรียน อีกทั้งเป็นข้อมูลเพื่อการปรับปรุงพัฒนาการเรียนการสอน ข้อมูลระดับสถานศึกษา สามารถนำไปใช้เพื่อวางแผนพัฒนาคุณภาพการจัดการศึกษาของสถานศึกษา แสดงศักยภาพของสถานศึกษาในการจัดการศึกษา รายงานต่อสาธารณชน/ชุมชนคณะกรรมการสถานศึกษาเพื่อแสดงความรับผิดชอบในการจัดการศึกษา เป็นข้อมูลเพื่อการปรับปรุงพัฒนาหลักสูตร การเรียนการสอน รวมทั้งใช้เป็นข้อมูลสร้างความเข้าใจความตระหนักในการมีส่วนร่วมพัฒนาคุณภาพของผู้เกี่ยวข้องสำหรับข้อมูลระดับเขตพื้นที่การศึกษา ระดับประเทศ หรือสังกัด สามารถนำไปใช้เพื่อการบริหารจัดการ กำหนดยุทธศาสตร์เพื่อการพัฒนาคุณภาพ ปรับเป้าหมาย นโยบาย ใช้เป็นข้อมูลพื้นฐานสำหรับการวางแผนช่วยเหลือ สนับสนุน จัดบุคลากร และดูระดับผลสัมฤทธิ์ในด้านต่าง ๆ โดยรวม เทียบกับคุณภาพตามมาตรฐานได้ (บุญชู ชลัษเฐียร, 2550, หน้า 2) สำนักทดสอบทางการศึกษา ได้ให้ความสำคัญกับการทดสอบวัดสัมฤทธิ์ระดับชาติ ดังนี้

1. ทำให้สามารถเปรียบเทียบผลการประเมินคุณภาพระหว่างชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ ตลอดจนการประเมินภายนอกอย่างสมเหตุสมผล

2. เพื่อกำกับ ติดตาม และควบคุม คุณภาพการจัดการศึกษาขั้นพื้นฐานของประเทศในช่วงชั้นที่ 1 (ประถมศึกษาปีที่ 3) และช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 3) เพื่อให้เกิดการพัฒนาอย่างต่อเนื่อง ซึ่งเป็นส่วนหนึ่งของการประกันคุณภาพ

3. เพื่อให้ได้ข้อมูลย้อนกลับ สำหรับกระบวนการจัดสินใจ และกำหนดแผนพัฒนาคุณภาพการจัดการศึกษาขั้นพื้นฐานของประเทศ เขตพื้นที่การศึกษา และระดับสถานศึกษา

4. สามารถประเมินได้ทั้งผลสัมฤทธิ์ทางวิชาการตามหลักสูตรและความถนัดทางการเรียน (Scholastic Aptitude Test: SAT)

5. ส่งเสริมและกระตุ้นให้สถานศึกษาให้ความสนใจอย่างจริงจังในการพัฒนาผลสัมฤทธิ์ที่สำคัญของหลักสูตร

6. สามารถใช้ผลการประเมินให้เป็นประโยชน์ทั้งในระดับผู้เรียน ระดับชั้นเรียนระดับสถานศึกษา ระดับเขตพื้นที่การศึกษาและระดับชาติ

7. สร้างแรงจูงใจกระตุ้นและทำนายให้ผู้เรียนทุกคนตั้งใจใฝ่หาสัมฤทธิ์ผลการเรียนและด้านอื่น ๆ

8. เพื่อเป็นข้อมูลสร้างความมั่นใจเกี่ยวกับคุณภาพผู้เรียนผู้เกี่ยวข้องทั้งภายในและภายนอกสถานศึกษา (นิพล พลกลาง, 2549, หน้า 39 – 40)

ตอนที่ 7 โปรแกรมคอมพิวเตอร์ที่ใช้ในการวิจัย

โปรแกรม HLM

นักวิธีวิทยาการวิจัยหลายท่านได้เสนอเทคนิคการประมาณค่าพารามิเตอร์ ตลอดจนโปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์ข้อมูลพหุระดับหลายวิธี เช่น Aitkin และ Longford นำเสนอโปรแกรม VARCL, Goldstien นำเสนอโปรแกรม ML/ 3, Bryk และ Raudenbush นำเสนอโปรแกรม HLM สำหรับวิธีการประมาณค่าพารามิเตอร์ที่สำคัญของการวิเคราะห์พหุระดับ ได้แก่ การวิเคราะห์ประมาณค่าส่วนประกอบความแปรปรวน (Analysis of Variance Component Estimation) วิธีกำลังสองน้อยที่สุดแบบสมการเดียว (OLS Separate Equation Approach) วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood) การประมาณค่าด้วยวิธีของเบส์ (Bayesian Estimation) เป็นต้นในปัจจุบันนอกจากการวิเคราะห์พหุระดับจะสามารถวิเคราะห์โมเดลที่มีตัวแปรตามตัวเดียว (Univariate Model) แล้ว ยังได้รับการพัฒนาให้สามารถวิเคราะห์โมเดลที่มีตัวแปรตามหลายตัว (Multivariate Model) ได้ในหลายลักษณะ เช่น การวิเคราะห์องค์ประกอบพหุระดับ การวิเคราะห์โค้งพัฒนาการชนิดตัวแปรแฝง การวิเคราะห์ถดถอยพหุระดับชนิดตัวแปรพหุ และการวิเคราะห์โมเดลสมการโครงสร้างพหุระดับ ซึ่งในการวิจัยครั้งนี้ผู้วิจัยได้ประยุกต์ใช้โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) ด้วยการวิเคราะห์ 2 ระดับ โดยการวิเคราะห์ด้วยโปรแกรม HLM

โปรแกรม Mplus

Muthén & Muthén (2007, 2010) พัฒนาโปรแกรม Mplus เพื่อให้นักวิจัยมีเครื่องมือสำหรับวิเคราะห์ข้อมูลด้วยสถิติวิเคราะห์ขั้นสูง ที่ให้ผลการวิเคราะห์ข้อมูลที่มีความถูกต้องมากกว่าสถิติวิเคราะห์แบบเดิม การพัฒนาโปรแกรม Mplus ดำเนินการเป็นกระบวนการที่มีการพัฒนาปรับปรุงโปรแกรมมาจนถึงปัจจุบัน โดยมีการเผยแพร่ โปรแกรม Mplus version 1 ปี 1998, version 2 ปี 2001, version 3 ปี 2004, version 4 ปี 2006, version 5 ปี 2007 และ version 5 ปี 2010 โปรแกรม Mplus version ใหม่ล่าสุด คือ โปรแกรม Mplus version 6.11

โปรแกรม Mplus ได้รับการพัฒนาให้เป็นโปรแกรมที่ใช้งานได้ง่ายและสะดวก และได้รับการปรับปรุงให้ดีขึ้น สามารถวิเคราะห์ข้อมูลได้หลายประเภท การใช้โปรแกรม Mplus จึงเป็นเรื่องง่ายและง่ายมากในกรณีที่นักวิจัยมีความรู้เรื่องการวิเคราะห์โมเดล SEM ด้วยโปรแกรม LISREL ความรู้พื้นฐานเรื่องโมเดล SEM จากคู่มือการใช้โปรแกรมของ Muthén & Muthén (2007, 2010) มี 4 หัวข้อ คือ 1) โมเดล SEM ในโปรแกรม Mplus 2) สัญลักษณ์และคำสั่งที่ใช้ในโปรแกรม 3) ภาษา Mplus และ 4) ตัวอย่างการใช้โปรแกรม Mplus ดังรายละเอียดต่อไปนี้

โมเดล SEM ในโปรแกรม Mplus

โมเดล SEM ในโปรแกรม Mplus แยกเป็น 4 ประเภท ตามลักษณะตัวแปร และลักษณะโมเดลดังนี้ นงลักษณ์ วิรัชชัย (2542)

1. โมเดลการวัดสาหรับตัวแปรแฝง – เมตริก หรือตัวแปรองค์ประกอบ (Factor)
2. โมเดลการวัดสาหรับตัวแปรแฝง – นันเมตริก หรือตัวแปรชั้นแฝง (Latent Class)
3. โมเดลสมการโครงสร้างที่มีตัวแปรภายในชนิดตัวแปรแฝง – เมตริก
4. โมเดลสมการโครงสร้างที่มีตัวแปรภายในชนิดตัวแปรแฝง – นันเมตริก

เมื่อรวมโมเดลทั้งสี่เข้าด้วยกันจะได้โมเดล SEM แบบต่าง ๆ ที่สามารถวิเคราะห์ข้อมูลได้โดยใช้โปรแกรม Mplus แยกเป็น 3 ประเภท ดังต่อไปนี้

1. โมเดล SEM แบบมีตัวแปรแฝง – เมตริก ได้แก่ โมเดลการวิเคราะห์ถดถอยพหุคูณ (Multiple Regression Analysis: MRA) การวิเคราะห์อิทธิพล (Path Analysis: PA) การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) โมเดลสมการโครงสร้าง (Structural Equation Model: SEM) โมเดลพัฒนาการมีตัวแปรแฝง (Latent Growth Model: LGM) และการวิเคราะห์การรอดชีพ (Survival Analysis: SV) ทั้งแบบช่วงเวลาต่อเนื่อง/ ไม่ต่อนื่อง

2. โมเดล SEM แบบมีตัวแปรแฝง – นันเมตริก ได้แก่ โมเดลการถดถอยพหุคูณแบบผสม (Multiple Regression Mixture Model: MRMM) โมเดลการวิเคราะห์อิทธิพลแบบผสม (Path Analysis Mixture Model: PAMM) การวิเคราะห์ชั้นแฝง (Latent Class Analysis: LCA) ทั้งแบบชั้นแฝงเดี่ยว/ ชั้นแฝงพหุ แบบมี/ ไม่มีตัวแปรเมตริก และแบบมี/ ไม่มีอิทธิพลทางตรง การวิเคราะห์ชั้นแฝงเชิงยืนยัน (Confirmatory Latent Class Analysis: CLCA) โมเดลล็อกลิเนียร์ (Loglinear Model) โมเดลนันพาราเมตริกของตัวแปรแฝง (Non – parametric Model of Latent Variable) การวิเคราะห์กลุ่มพหุ (Multiple Group Analysis) โมเดลความสัมพันธ์เชิงสาเหตุเฉลี่ย (Complier Average Causal Effect: CACE Model) การวิเคราะห์พัฒนาการชั้นแฝง (Latent Class Growth Analysis) และการวิเคราะห์การรอดชีพแบบผสม (Survival Mixture Analysis: SMV) ทั้งแบบช่วงเวลาต่อเนื่อง/ ไม่ต่อเนื่อง

3. โมเดล SEM แบบมีทั้งตัวแปรแฝง – เมตริก และตัวแปรแฝง – นันเมตริก ครอบคลุมโมเดลประเภทที่ 1 และ 2 ทุกโมเดล นอกจากนี้ยังครอบคลุมโมเดลต่อไปนี้ด้วย คือ โมเดลการวิเคราะห์ชั้นแฝงแบบมีอิทธิพลสุ่ม (Latent Class Analysis with Random Effects) โมเดลการวิเคราะห์องค์ประกอบแบบผสม (Factor Analysis Mixture Model) โมเดลสมการโครงสร้างแบบผสม (Structural Equation Mixture Model) และโมเดลพัฒนาการแบบผสมที่มีชั้นแฝงวิถีโค้ง (Growth Mixture Model with Latent Trajectory Class)

4. โมเดลที่มีข้อมูลจากการสำรวจซับซ้อน (Model with Complex Survey Data) ประกอบด้วยโมเดลพหุระดับ (Multilevel Model) ที่กลุ่มตัวอย่างได้มาจากการแบ่งชั้น (Cluster Sampling)

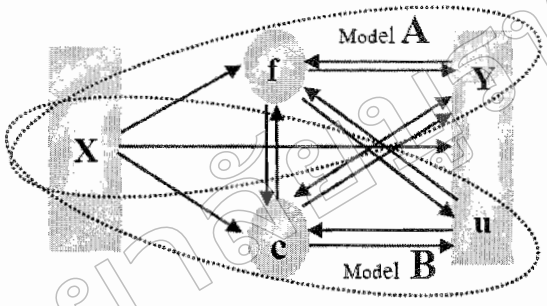
5. Muthén & Muthén (2007, 2010) สรุปประเภทโมเดลในโปรแกรม Mplus อีกแบบหนึ่ง โดยแยกประเภทตามลักษณะการวัดของตัวแปรในโมเดล แยกเป็น 3 โมเดล คือ

5.1 โมเดลกลุ่ม A ประกอบด้วย Multiple Regression Analysis (MRA), Analysis of Variances (ANOVA), Path Analysis (PA), Exploratory Factor Analysis (EFA), Structural Equation Modeling (SEM) (Including CFA, IRT, MIMIC (CFA with Covariates), MRA, MMRA, PA, Multiple Group Analysis (MG) and MG Analysis of the Above Models), Growth Modeling, Discrete – time Survival Analysis, Continuous – time Survival Analysis

5.2 โมเดลกลุ่ม B ประกอบด้วย Multiple Regression Mixture Modeling, Path Analysis Mixture Modeling, Latent Class Analysis, Confirmatory Latent Class Analysis, Latent Class Analysis with Multiple Categorical Latent Variables, Log Linear Modeling,

Non – parametric Modeling of Latent Variable Distribution, Multiple Group Analysis, Finite Mixture Modeling, Complier Average Causal Effect (CAFE) Modeling, Latent Transition Analysis & Hidden Markov Modeling, Latent Class Growth Analysis

5.3 โมเดลเต็มรูป (Full Models) ประกอบด้วยโมเดลทุกโมเดลในโมเดลกลุ่ม A และกลุ่ม B และโมเดลสมการโครงสร้างพหุระดับ (Multi – level SEM) ทั้งหมด แสดงในภาพที่ 2 – 5



ภาพที่ 2 – 13 โมเดล SEM ในโปรแกรม Mplus (Muthén & Muthén 2007, 2010)

X = Continuous Observed Exogenous Variables

Y = Continuous Observed Endogenous Variables

u = Categorical Observed Endogenous Variables

f = Continuous Latent Variables

c = Categorical Latent Variable

Models A = Models with only Continuous Latent Variables

Models B = Models with only Categorical Latent Variables

โปรแกรม WinBUGS

การวิเคราะห์ข้อมูลด้วยวิธีเบส์เซียน โดยใช้โปรแกรม WinBUGS

โมเดลเบส์เซียนและการอนุมานเบส์เซียน มีความแตกต่างจากข้อสรุปแบบดั้งเดิม การอนุมานรูปแบบความน่าจะเป็นของโมเดลสำหรับตัวแปรสังเกตและพารามิเตอร์ที่ไม่ทราบค่า เบส์เซียนจะอนุมานโดยการตรวจสอบเงื่อนไขของพารามิเตอร์ที่แสดงถึงข้อมูลที่สังเกตได้ ดังนั้น จึงสามารถใช้งานได้ง่าย ค่าโพสทีเรียที่ได้จะมีค่าพารามิเตอร์มากกว่าศูนย์ คือ .95 เพราะว่า ข้อตกลงของโมเดลเบส์เซียน มีการกระจายร่วมกันระหว่างข้อมูลและค่าพารามิเตอร์ โดยโมเดลเบส์เซียน เป็นวิธีการที่ยืดหยุ่น สามารถแก้ ปัญหาที่ง่ายและซับซ้อน มีการกำหนดค่าการแจกแจงเริ่มต้น ที่สามารถใช้ในการกำหนดช่วงของค่า พารามิเตอร์ของโมเดลพหุระดับ ความยืดหยุ่น

ของวิธีเบส์เซียน ทำให้เป็นประโยชน์อย่างยิ่งสำหรับการวิเคราะห์ข้อมูล Psychometric ความยากหนึ่งในการวิเคราะห์ข้อมูลโดยวิธีเบส์เซียนมีหลาย ๆ อย่าง แต่ไม่ใช่เรื่องง่าย ที่จะกำหนดการแจกแจงแบบโพสทีเรียในการวิเคราะห์ค่าพารามิเตอร์ของโมเดล แต่มากกว่านั้น โมเดลที่มีความซับซ้อน โพสทีเรียสามารถกำหนดค่าที่ใกล้เคียง แต่โมเดลที่ใกล้เคียงนี้สามารถ ทำงานได้ดีค่อนข้างน้อย การพัฒนาและปรับปรุงของ Monte Carlo จึงเป็นเทคนิคที่ได้ทำ เมื่อไม่นานมานี้ การแจกแจงแบบโพสทีเรียของเบส์เซียน มีความซับซ้อนมากขึ้น และซอฟแวร์ ที่ใช้งานง่ายและประมาณค่าได้ก็คือ โปรแกรม WinBugs ทำให้ผู้ใช้มีเครื่องมือที่ใช้งานง่าย โดยมีวิธีการนำเทคนิค Markov Chain Monte Carlo (MCMC) มาใช้

การใช้งานโปรแกรมใน WinBUGS สามารถทำได้ 2 วิธี คือ

1. การจำลองข้อมูล (Simulation) จากโปรแกรม R แล้วทำการประมวลผลภายใต้ จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS ด้วย Package R2 WinBUGS

2. การเขียนคำสั่งในโปรแกรม WinBUGS จากนั้นทำการประมวลผลด้วยโปรแกรม WinBUGS ได้เลย

ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

งานวิจัยต่างประเทศ

Swaminathan & Rogers (1990) ได้เปรียบเทียบระหว่างวิธีการถดถอยโลจิสติก และ วิธีแมนเทล – แฮนส์เซล ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและ แบบอเนกรูป โดยศึกษาในสถานการณ์จำลอง 6 เงื่อนไข คือ ขนาดกลุ่มตัวอย่าง 2 ระดับ (250 คน ต่อกลุ่ม และ 500 คนต่อกลุ่ม) และความยาวของแบบสอบ 3 ระดับ (40 ข้อ 60 ข้อ และ 80 ข้อ) ซึ่งแบบสอบแต่ละชุดประกอบด้วยสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันจำนวน 20% โดยครึ่งหนึ่ง เป็นข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป และอีกครึ่งหนึ่งเป็นข้อสอบที่ทำหน้าที่ต่างกัน แบบอเนกรูปสำหรับผลการตอบข้อสอบทั้งหมดจำลองโดยใช้โปรแกรม DATAGEN โมเดล การตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ในการจำลองข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จะกำหนดให้พารามิเตอร์อำนาจจำแนกระหว่างผู้สอบ 2 กลุ่มมีค่าเท่ากัน ในขณะที่พารามิเตอร์ ความยากจะมีค่าแปรเปลี่ยนสำหรับการจำลองข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จะกำหนดให้พารามิเตอร์ความยากระหว่างผู้สอบ 2 กลุ่มมีค่าเท่ากัน ส่วนพารามิเตอร์ อำนาจจำแนกจะมีค่าแปรเปลี่ยนในการควบคุมขนาดของการทำหน้าที่ต่างกันของข้อสอบ จะใช้พื้นที่ระหว่างโค้งลักษณะข้อสอบ ซึ่งคำนวณโดยใช้สูตรของ Raju

ผลการศึกษา พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป วิธีการถดถอยโลจิสติกมีอำนาจการทดสอบสูงกว่าวิธีแมนเทล – แฮนส์เซล ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปทั้ง 2 วิธีมีอำนาจการทดสอบเท่าเทียมกัน สำหรับปัจจัยของขนาดกลุ่มตัวอย่าง พบว่า เมื่อขนาดกลุ่มตัวอย่าง 250 คนต่อกลุ่มผู้สอบ ทั้ง 2 วิธีมีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปประมาณ 75% และเมื่อขนาดกลุ่มตัวอย่าง 500 คนต่อกลุ่มผู้สอบ ทั้ง 2 วิธีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปประมาณ 100% สำหรับอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 พบว่า วิธีแมนเทล – แฮนส์เซล มีอัตราความคลาดเคลื่อนร้อยละ 1 ในขณะที่วิธีการถดถอยโลจิสติกมีอัตราความคลาดเคลื่อนระหว่างร้อยละ 1 ถึง 6 เมื่อพิจารณาถึงค่าใช้จ่ายในการวิเคราะห์ ปรากฏว่าวิธีการถดถอยโลจิสติกมีค่าใช้จ่ายสูงกว่าวิธีแมนเทล – แฮนส์เซล ประมาณ 3 – 4 เท่า

Raju, Drasgow, & Slinde (1993) ได้ศึกษาเปรียบเทียบการประเมินการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวัดพื้นที่ (ชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมาย) วิธีการทดสอบไค – สแควร์ของ Lord และวิธีแมนเทล – แฮนส์เซล โดยศึกษากับข้อมูลเชิงประจักษ์ ใช้กลุ่มตัวอย่างนักเรียนระดับ 10 และระดับ 12 จำนวน 839 คน การแบ่งกลุ่มตัวอย่างเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบใช้สีผิวและเพศเป็นเกณฑ์ ในกรณีใช้สีผิวเป็นเกณฑ์ได้นักเรียนผิวดำ 245 คน และผิวขาว 436 คน ส่วนที่เหลืออีก 158 คน แตกต่างไปจากทั้งสองกลุ่มจึงคัดออก ในกรณีใช้เพศเป็นเกณฑ์ได้นักเรียนหญิง 440 คน และนักเรียนชาย 399 คน ข้อมูลที่ใช้ในการศึกษาได้มาจากการทดลองสอบกับนักเรียนระดับ 10 และ 12 ในปี ค.ศ.1987 โดยใช้แบบสอบ Gates – MacGinitie Reading Tests (GMRT) ซึ่งเป็นชุดข้อสอบที่ใช้วัดความรู้เกี่ยวกับคำศัพท์จำนวน 45 ข้อ ในแต่ละข้อมี 5 ตัวเลือก สำหรับในการวิเคราะห์ข้อมูลจะเริ่มต้นด้วยการตรวจสอบความเป็นเอกมิติของแบบสอบโดยการวิเคราะห์องค์ประกอบหลักกับผลการตอบข้อสอบของกลุ่มตัวอย่างทั้งหมด แล้วจึงประมาณค่าพารามิเตอร์ของข้อสอบโมเดลโลจิสติกแบบ 2PLM โดยใช้โปรแกรม BILOG ด้วยวิธีประมาณค่าของ Bayes ซึ่งจะประมาณค่าพารามิเตอร์แยกกลุ่มผู้สอบออกเป็น 4 กลุ่ม คือ กลุ่มผิวดำ กลุ่มผิวขาว กลุ่มหญิง และกลุ่มชาย ต่อจากนั้นจึงใช้โปรแกรม EQUATE ด้วยวิธีเทียบโค้งลักษณะข้อสอบของ Stocking และ Lord เพื่อแปลงค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบให้อยู่บนสเกลเดียวกัน หลังจากนั้นจึงคำนวณการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มผู้สอบดำ – ขาว และกลุ่มผู้สอบชาย – หญิง โดยใช้วิธีการวัดพื้นที่ชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมายของ Raju

วิธีการทดสอบไค – สแควร์ของ Lord และวิธีแมนเทล – แฮนส์เซล สำหรับการทดสอบนัยสำคัญทางสถิติกับวิธีแรกใช้สถิติ Z ส่วนสองวิธีหลังใช้สถิติไค – สแควร์

ผลการศึกษา พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกรณีการแบ่งกลุ่มตามเพศและสีผิว วิธีการวัดพื้นที่ชนิดคิดเครื่องหมาย วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมาย และวิธีการทดสอบไค – สแควร์ของ Lord สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้สอดคล้องกันอย่างมีนัยสำคัญ เฉพาะกรณีการแบ่งกลุ่มตามเพศ วิธีการวัดพื้นที่ชนิดคิดเครื่องหมาย การวัดพื้นที่ชนิดไม่คิดเครื่องหมาย วิธีการทดสอบไค – สแควร์ของ Lord และวิธีแมนเทล – แฮนส์เซล สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้สอดคล้องกันสูง ส่วนในกรณีการแบ่งกลุ่มตามสีผิวจะระบุข้อสอบที่ทำหน้าที่ต่างกันได้แตกต่างกัน

Budgell, Raju, & Quartetti (1995) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในเครื่องมือการประเมินที่แปลเป็นสองภาษา โดยใช้วิธีการตรวจสอบ 2 วิธี 1) วิธีการวัดพื้นที่ชนิดคิดเครื่องหมาย (LC) และ 2) วิธีแมนเทล – แฮนส์เซล (MH) วิธีการวัดพื้นที่ทั้งชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมายเป็นวิธีวัดพื้นที่ในช่วงเปิดของ Raju เครื่องมือที่ใช้ในการศึกษาเป็นแบบสอบชนิดเลือกตอบ ซึ่งวัดด้านตัวเลขจำนวน 15 ข้อ และวัดด้านเหตุผลจำนวน 18 ข้อโดยแยกเป็น 2 ฉบับ คือฉบับภาษาอังกฤษและฉบับภาษาฝรั่งเศส แบบสอบทั้ง 2 ฉบับพัฒนามาจากชุดแบบสอบวัดความรู้ความสามารถทั่วไปที่ใช้ในประเทศแคนาดาโดยคณะผู้เชี่ยวชาญของการแปลภาษาทั้งสอง สำหรับการดำเนินการสอบจะต้องควบคุมให้เป็นไปตามเงื่อนไข 2 ประการ คือ ประการแรกผู้สอบต้องเลือกแบบสอบให้ตรงกับภาษาที่ใช้มาตั้งแต่กำเนิด และประการที่สองผู้สอบที่อาศัยอยู่ในประเทศแคนาดาจะต้องเลือกแบบสอบให้ตรงกับภาษาหลักที่ใช้กันในชุมชน มีผู้เข้าสอบทั้งหมด 16,362 คน ต่อจากนั้นจะสุ่มกลุ่มตัวอย่างมา 4 กลุ่ม ๆ ละ 1,000 คน กลุ่มผู้สอบรูปแบบภาษาอังกฤษ 2 กลุ่ม (E_1 และ E_2) และกลุ่มผู้สอบรูปแบบภาษาฝรั่งเศส 2 กลุ่ม (F_1 และ F_2) นำผู้สอบทั้ง 4 กลุ่มไปจับคู่เปรียบเทียบเพื่อวิเคราะห์หาดัชนี DIF โดยเปรียบเทียบกลุ่มผู้สอบ 4 เงื่อนไข คือ 1) E_1 กับ $F_1$2) E_2 กับ F_2 3) E_1 กับ E_2 และ 4) F_1 กับ F_2 ในการวิเคราะห์แบบสอบทั้ง 2 ฉบับจะวิเคราะห์แยกกันทั้งรูปแบบภาษา (อังกฤษและภาษาฝรั่งเศส) และเนื้อหาที่ใช้วัด (ตัวเลขและเหตุผล) นำแบบสอบทั้ง 2 ฉบับไปวิเคราะห์องค์ประกอบโดยใช้การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) เพื่อตรวจสอบความเป็นเอกมิติก่อนที่จะนำไปวิเคราะห์ตามทฤษฎี IRT ใช้โปรแกรม BILOG ประมาณค่าพารามิเตอร์ของข้อสอบตามโมเดลโลจิสติกแบบ 2 พารามิเตอร์ และใช้โปรแกรม

EQUATE เทียบค่าพารามิเตอร์ของข้อสอบจากผู้สอบกลุ่มย่อย 2 กลุ่ม สำหรับสถิติที่ใช้ทดสอบดัชนี DIF ประกอบด้วย สถิติ Z ทดสอบดัชนี SA และ UA ($Z > -2.58$ หรือ $Z > +2.58$ เมื่อ $\alpha = .01$) สถิติ χ^2 ทดสอบดัชนี LC (χ^2 เท่ากับ 9.21 เมื่อ $\alpha = .01$, $df = 2$) และสถิติ χ^2 ทดสอบดัชนี MH (χ^2 เท่ากับ 6.63 เมื่อ $\alpha = .01$, $df = 1$)

ผลการศึกษา พบว่า การเปรียบเทียบตามเงื่อนไข 1 และ 2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SA, UA, LC และ MH สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้สอดคล้องกันอย่างมีนัยสำคัญ โดยเฉพาะในแบบสอบวัดด้านตัวเลขและแบบสอบวัดด้านเหตุผลทั้ง 4 วิธี สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันได้สอดคล้องกันสูง ยกเว้นการตรวจสอบด้วยวิธี UA ในแบบสอบวัดด้านเหตุผล สามารถระบุข้อสอบที่ทำหน้าที่ต่างกันต่ำกว่าการตรวจสอบด้วยวิธีอื่น ๆ สำหรับการเปรียบเทียบตามเงื่อนไข 3 และ 4 ปรากฏว่าไม่พบข้อสอบที่ทำหน้าที่ต่างกัน

Welch & Miller (1995, pp. 163 – 178) ได้ศึกษาเปรียบเทียบผลของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า 3 วิธี คือ วิธีการ HW3 วิธีการ GMH และวิธีการ LDFA ภายใต้เงื่อนไขของการรวมกันของตัวแปรจับคู่ภายในและภายนอกที่ต่างกัน 3 เงื่อนไข คือ เงื่อนไข A เป็นการจับคู่โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบจำนวน 40 ข้อ เงื่อนไข B เป็นการจับคู่โดยใช้โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบจำนวน 40 ข้อกับคะแนนจากข้อสอบประเมินการเขียนแบบอธิบายความ เงื่อนไข C เป็นการจับคู่โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบ จำนวน 40 ข้อกับคะแนนจากข้อสอบประเมินการเขียนตอบแบบอธิบายความและแบบพรรณนาความ กลุ่มตัวอย่างเป็นนักเรียนเกรด 8 แบ่งเป็นกลุ่มย่อย 4 กลุ่ม คือ กลุ่มอเมริกันผิวดำ กลุ่มอเมริกันผิวขาว กลุ่มเพศชายและกลุ่มเพศหญิง เครื่องมือที่ใช้เป็นแบบทดสอบประเมินการเขียนตอบ แบ่งเป็น 3 รูปแบบ คือ 01A 01B และ 01C โดยรูปแบบ 01A มีผู้สอบจำนวน 1,497 คน รูปแบบ 01B มีผู้สอบจำนวน 1,523 คน และรูปแบบ 01C มีผู้สอบจำนวน 1,233 คน สัมประสิทธิ์การอ้างอิงของแบบทดสอบการประเมินการเขียนตอบเป็น .60 ส่วนสัมประสิทธิ์ความเชื่อมั่นแบบความสอดคล้องภายใน (แอลฟา) ของข้อสอบแบบเลือกตอบ เป็น .09

ผลการศึกษา พบว่า ผลการประเมินการเขียนจากแบบทดสอบการเขียนและแบบทดสอบเลือกตอบเพศหญิงทำได้ดีกว่าเพศชาย และอเมริกันผิวขาวทำได้ดีกว่าอเมริกันผิวดำ ซึ่งผลเช่นนี้ยังไม่สามารถสรุปได้ว่ามี DIF เสมอไป ต้องทำการวิเคราะห์ต่อโดยใช้วิธีการที่สามารถแยกผลกระทบ (Impact) ออกจากการทำหน้าที่ต่างกันจริง (True DIF) ส่วนค่าสหสัมพันธ์ระหว่าง

ข้อสอบประเมินการเขียนตอบกับข้อสอบแบบเลือกตอบมีความสอดคล้องกันอยู่ในช่วง .66 ถึง .77 ซึ่งบ่งชี้ว่ามีความแตกต่างในการวัดด้วยเครื่องมือวัด 2 ชนิดนี้ แต่เมื่อรวมข้อสอบประเมินการเขียนตอบเข้ากับข้อสอบแบบเลือกตอบ ค่าสหสัมพันธ์ยิ่งมีค่าสูงขึ้น

Rijiman, Tuerlinckx, Meulders, Smits, & Balazs. (2005) ได้วิเคราะห์วิธีการประมาณค่าโมเดลผสมสำหรับโมเดลราสช์ เพื่อประเมินผลของการประมาณค่าโมเดลผสมสำหรับโมเดลเส้นตรงทั่วไป และโมเดลผสมไม่เป็นเส้นตรง ซึ่ง Rijman และคณะได้กล่าวว่า โมเดลผสม (Mixed models) เป็นชุดของเครื่องมือทางสถิติที่มีความเหมาะสมสำหรับการวิเคราะห์ข้อมูลที่มีลักษณะเป็นกลุ่ม ๆ และสอดแทรกอยู่ด้วยกัน ในการศึกษาครั้งนี้ มีการจำลองข้อมูล เพื่อศึกษาลักษณะการประมาณค่าที่ต่างกันของโมเดลราสช์ตามเงื่อนไข ดังนี้ กลุ่มคนมี 2 กลุ่ม คือ 100 คน และ 500 คน จำนวนข้อสอบ มี 2 กลุ่ม คือ 5 ข้อ และ 25 ข้อ และค่าพารามิเตอร์ของข้อสอบมีค่าระหว่าง – 2 ถึง 2 โดยที่ในการวิเคราะห์แต่ละครั้งจะประกอบด้วยข้อมูลจำนวน 30 ชุด ทั้ง 8 เงื่อนไข ซึ่งของการประเมินความเหมาะสมของการประมาณค่าพารามิเตอร์นั้น Rijman และคณะได้สร้างดัชนีสำหรับตรวจสอบและประเมินความสามารถในการประมาณค่า (GOR: Goodness of Recovery) จำนวน 3 ดัชนี ดังนี้ 1) BIAS: Different Between the Average Estimated Parameter 2) RMSD: Root Mean Square Deviation 3) MCSE: Monte Carlo Standard Error โดยดัชนีทั้ง 3 มีความสัมพันธ์กัน คือ $RMSD^2 = MCSE^2 + BIAS^2$

การประมาณค่าพารามิเตอร์จะดำเนินการผ่านวิธีการคำนวณ 4 วิธีการ ดังนี้ 1) วิธีการ Gaussian Quarture โดยใช้โปรแกรม SAS: NLMIXED 2) วิธีการ Sixth – order Laplace โดยใช้โปรแกรม HLM (V.5.04) 3) วิธีการ PQL2 โดยใช้โปรแกรม MLwiN (V.1.10) และ 4) วิธีการ MCMC โดยใช้โปรแกรม WinBUGS (V1.2)

Finch (2005) เปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธีแมนเทลเฮนเซล (Mantel & Haenszel, 1959) และวิธี SIBTEST (Shealy & Stout, 1993) และวิธีการทดสอบ IRT Likelihood Ratio (Thissen et al., 1986) กับความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้แสดงให้เห็นว่าวิธี MIMIC มีค่าสูงขึ้นและความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงกับจำนวนข้อสอบ 50 ข้อ นอกจากนี้วิธี MIMIC ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบ Uniform DIF ได้ด้วย และ Fukahara & Kamata (2007) ได้มีการประเมินประสิทธิภาพการทำงานของวิธี MIMIC แบบละเมิดข้อตกลงเบื้องต้น โดยการทำชุดข้อสอบ พบว่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) มีแนวโน้มที่จะประเมินค่าต่ำกว่า เมื่อข้อสอบไม่เป็นอิสระ

งานวิจัยในประเทศ

วิชุดา บัวคง (2533) ได้เปรียบเทียบประสิทธิผลของวิธีการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้เข้าสอบระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติค และวิธีแบส์ ของแบบจำลองโลจิสติค 3 พารามิเตอร์ โดยศึกษาจากแบบทดสอบวัดผลสัมฤทธิ์วิชาภาษาไทย (ด้านการใช้ภาษา) และแบบทดสอบความถนัดด้านการใช้ภาษาไทย โดยการประมาณค่าพารามิเตอร์ความยาก อำนาจจำแนก การเดาของข้อสอบ และความสามารถของผู้เข้าสอบตามวิธีประมาณค่าทั้ง 3 วิธี คำนวณค่าฟังก์ชั้นสารสนเทศของข้อสอบและแบบทดสอบ ค่าประสิทธิภาพสัมพัทธ์และความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ของแบบทดสอบ

ผลการวิจัย พบว่า แบบทดสอบวัดผลสัมฤทธิ์ที่ประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีแมกซิมัม ไลค์ลิฮูด มีประสิทธิภาพสูงที่สุด ในกลุ่มผู้สอบที่มีความสามารถสูง รองลงไป คือ แบบทดสอบที่ประมาณค่าด้วยวิธีของเบส์ และวิธีฮิวริสติค ตามลำดับ ส่วนในกลุ่มผู้เข้าสอบที่มีความสามารถปานกลางและต่ำ แบบทดสอบที่ประมาณค่าด้วยวิธีของเบส์ มีประสิทธิภาพสูงที่สุด รองลงไป คือ แบบทดสอบที่ประมาณค่าด้วยวิธีแมกซิมัม ไลค์ลิฮูด และวิธีฮิวริสติค สำหรับค่าความเที่ยงตรงตามสภาพที่เป็นผลจากการประมาณค่าความสามารถของผู้เข้าสอบทั้ง 3 วิธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 โดยวิธีของเบส์ ให้ค่าความเที่ยงตรงตามสภาพสูงที่สุด รองลงไป คือ วิธีฮิวริสติค หรือวิธีแมกซิมัม ไลค์ลิฮูด

กาญจนา วัธนสุนทร (2537) ได้พัฒนาเกณฑ์ได้การตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ โดยวิธีการตรวจสอบ 4 วิธี คือ วิธีการวัดพื้นที่ชนิดคิดเครื่องหมาย (SA) วิธีการวัดพื้นที่ชนิดไม่คิดเครื่องหมาย (UA) วิธีแมนเทล – แฮนส์เซล (MH) และวิธีซิปเทสท์ (SIBTEST) ข้อมูลที่ใช้ในการศึกษาครั้งนี้ได้จากการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 โดยศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 2 ตัว คือ ความยาวของแบบสอบ ซึ่งมี 2 วิชา ได้แก่ แบบสอบวิชาคณิตศาสตร์ 3 ระดับ (20 ข้อ 30 ข้อ และ 40 ข้อ) แบบสอบวิชาภาษาอังกฤษ 4 ระดับ (50 ข้อ 60 ข้อ 70 ข้อ และ 80 ข้อ) และขนาดกลุ่มตัวอย่าง 6 ระดับ (100 คน 200 คน 400 คน 600 คน 800 คน และ 1,000 คน) ในการประมาณค่าพารามิเตอร์ของข้อสอบโปรแกรม BILOG โมเดลโลจิสติกแบบ 2 PLM โดยใช้วิธีการประมาณค่าแบบ MMLE แล้วใช้โปรแกรม EQUATE ปรับเทียบสเกลพารามิเตอร์ของข้อสอบที่ประมาณค่าจากผู้สอบสองกลุ่ม คือ กลุ่มอ้างอิงและกลุ่มเปรียบเทียบซึ่งจำแนกตามเพศ ในการคำนวณดัชนี DIF ด้วยวิธีการวัดพื้นที่ชนิดคิดเครื่องหมายและชนิดไม่คิดเครื่องหมายใช้โปรแกรม AREA ส่วนวิธี MH และวิธี SIBTEST ในการพัฒนาดัชนี

เพื่อใช้เป็นเกณฑ์สำหรับการตัดสินความลำเอียงของข้อสอบมี 4 ตัว คือ SA, UA, α_{MH} และ β_{SIB} โดยจะวิเคราะห์ค่าเฉลี่ยของดัชนีแต่ละตัวแล้วกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะ คือ เกณฑ์ที่กำหนดจากค่าเฉลี่ยซึ่งรวมค่าดัชนีทุกข้อโดยไม่ได้พิจารณาถึงปัจจัยความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง ส่วนเกณฑ์อีกลักษณะหนึ่งจะกำหนดจากค่าเฉลี่ยซึ่งได้พิจารณาถึงปัจจัยความยาวของแบบสอบและขนาดของกลุ่มตัวอย่าง แล้วจึงนำเกณฑ์ที่กำหนดไว้ไปตัดสินค่าดัชนีที่ได้จากการวิเคราะห์ระหว่างกลุ่มผู้สอบเพศชายและหญิง

ผลการวิจัย พบว่า ขนาดกลุ่มตัวอย่างมีผลต่อค่าเฉลี่ยของดัชนีทุกตัว ส่วนความยาวของแบบสอบมีผลต่อค่าเฉลี่ยของดัชนี SA และ UA แต่ไม่มีอิทธิพลต่อค่าเฉลี่ยของดัชนี α_{MH} และ β_{SIB} สำหรับเกณฑ์ที่พัฒนาเพื่อให้ตัดสินความลำเอียงของข้อสอบระหว่างผู้สอบเพศชายและเพศหญิงเป็น ดังนี้

1. $|SA| > .80$ และ $UA > .50$ เมื่อความยาวของแบบสอบน้อยกว่า 50 ข้อ
2. $|SA| > .40$ และ $UA > 1.20$ เมื่อความยาวของแบบสอบมากกว่า 50 ข้อ
3. $\alpha_{MH} < .60$, $\alpha_{MH} > 1.40$ และ $|\beta_{SIB}| > .06$ สำหรับทุกความยาวของแบบสอบและขนาดกลุ่มตัวอย่าง

เกสร หว่างจิตร (2539) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทลแฮนส์เซล โดยใช้เพศ ภูมิลำเนา ประสบการณ์ในการสอบ และสังกัดของสถานศึกษา เป็นเกณฑ์จำแนกกลุ่มผู้สอบเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบในวิชาสอบร่วมของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในส่วนเฉพาะที่เป็นข้อสอบชนิดเลือกตอบ 2 วิชา คือ วิชาภาษาไทยซึ่งมีผู้สอบจำนวน 506 คน และวิชาภาษาอังกฤษซึ่งมีผู้สอบจำนวน 501 คน แล้วนำผลการสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปโดยโปรแกรม MH_{DIF} สำหรับการตรวจสอบความเที่ยงและความตรงของแบบสอบใช้โปรแกรม CTIA และ LISREL ตามลำดับ

ผลการศึกษา พบว่า ข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันส่วนมากมีลักษณะเป็นแบบอเนกรูป โดยพบในกลุ่มผู้สอบย่อยที่จำแนกตามเพศมากที่สุด รองลงมาคือ จำแนกตามภูมิลำเนา สังกัดของสถานศึกษา และประสบการณ์ในการสอบตามลำดับ ข้อสอบที่ทำหน้าที่ต่างกันส่วนมากเป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำทั้งสองวิชา เมื่อพิจารณาลักษณะของข้อสอบพบว่า ในแบบสอบวิชาภาษาไทยข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่ง่ายมากแต่ในแบบสอบวิชาภาษาอังกฤษข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันส่วนมากจะเป็นข้อสอบที่ยากมาก นอกจากนี้ยังพบว่า ค่าความเที่ยงและความตรงของแบบสอบก่อนและหลังการตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบสอบไม่แตกต่างกันทั้งสองวิชา

เรวดี อินทะสะระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศ ด้วยวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) วิธีแมลเทล – แฮนส์เซล (MH) และวิธีซิปเทสท์ (SIBTEST) พร้อมทั้งศึกษาการตัดสินผลการสอบที่คิดคะแนนมาตรฐานทีปกติ และคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาเอกภาษาไทย ก วิชาสังคม ก และวิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (เป็นชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน (เป็นชาย 1,451 คน หญิง 3,691 คน)

ผลการศึกษา พบว่า วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคม ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นแตกต่างกันที่ระดับนัยสำคัญ .01 โดยวิธีทฤษฎีการตอบข้อสอบตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุดความสัมพันธ์ของลำดับที่ของการตอบ ไม่ว่าจะคิดคะแนนมาตรฐานปกติที หรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมด หรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

พรรณี จิตมาศ (2540) ได้วิเคราะห์ความลำเอียงต่อเพศของแบบทดสอบคณิตศาสตร์โจทย์ปัญหาที่ผู้วิจัยสร้างขึ้นเอง ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธี MH และวิธี SIBTEST ในแต่ละขนาดของกลุ่มผู้สอบ 500 และ 1000 คน โดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลางจำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดของโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยสุ่ม เครื่องมือที่ใช้คือแบบทดสอบคณิตศาสตร์โจทย์ปัญหาเรื่องสมการ อัตราส่วน และร้อยละ เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง จำนวน 1 ฉบับ

ผลการวิจัย พบว่า เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธี SIBTEST พบข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทุกวิธีวิเคราะห์ และเมื่อวิเคราะห์จากลุ่มผู้สอบขนาด 1,000 คน วิธี MH พบข้อสอบมีความลำเอียงมากที่สุด

ญาณภัทร สีหะมงคล (2540) เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพ

และความก้าวหน้าทางการศึกษา วิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 ของสำนักงานการประถมศึกษาแห่งชาติ จำนวน 11,404 คน เครื่องมือที่ใช้เป็นแบบทดสอบแบบเลือกตอบ จำนวน 80 ข้อ

ผลการศึกษา พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสามแตกต่างกัน เมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ส่วนความ สัมพันธ์ระหว่างวิธีการทั้งสามมีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษาและสำหรับความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนมากจะมีค่าปานกลางถึงต่ำ เกือบทุกเงื่อนไขของการศึกษา

วลีมาศ แซ่อึ้ง (2543) ทำการเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล – แฮนส์เซลและวิธีถดถอยโลจิสติก โดยมีเงื่อนไขที่ทำ การศึกษา 324 เงื่อนไข

ผลการศึกษา พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิปเทสท์ปรับใหม่ และวิธีถดถอยโลจิสติกมีค่าเท่าเทียมกันในทุกเงื่อนไข และทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิปเทสท์และวิธีแมนเทล – แฮนส์เซลภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอเนกรูปของวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล – แฮนส์เซล และวิธีถดถอยโลจิสติกมีค่าอยู่ในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ระดับ 10% ทุกเงื่อนไข

รักชนก ยิ่สุนศรี (2544) ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษา เพื่อเปรียบเทียบความเที่ยง ความตรง และฟังก์ชันสารสนเทศของแบบสอบระหว่างแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF โดยใช้ข้อมูลจากการตอบแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและคณิตศาสตร์ ปีการศึกษา 2543 ครั้งที่ 1 และเลือกศึกษาในส่วนที่เป็นข้อสอบแบบหลายตัวเลือกจากกลุ่มตัวอย่างที่เป็นผู้เข้าสอบจำนวน 4,000 คน และ 3,600 คน ตามลำดับ

ผลการวิจัยพบว่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้ดัชนี NCDIF พบว่าแบบสอบวิชาภาษาอังกฤษ เมื่อจำแนกกลุ่มผู้สอบตามเพศ มีข้อสอบที่พบ DIF จำนวน 30 ข้อ คิดเป็นร้อยละ 30.00 ของจำนวนข้อสอบทั้งหมด และเมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีข้อสอบที่พบ DIF จำนวน 16 ข้อ คิดเป็นร้อยละ

16.00 ของจำนวนข้อสอบทั้งหมด นอกจากนี้ในแบบสอบวิชาคณิตศาสตร์ เมื่อจำแนกกลุ่มผู้สอบตามสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษา มีข้อสอบที่พบ DIF จำนวน 16 ข้อ คิดเป็นร้อยละ 21.43 ของจำนวนข้อสอบทั้งหมด

นพดล มีชั้นช่วง (2544) ได้ศึกษาการเปรียบเทียบผลของการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ ระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติก และวิธีของเบย์ของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1

ผลการวิจัย พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยากที่ประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติก และวิธีของเบย์ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าตั้งแต่ .263 ถึง .446 วิธีแมกซิมัมไลค์ลิฮูด และวิธีของเบย์ มีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ วิธีแมกซิมัมไลค์ลิฮูด กับวิธีของเบย์ ส่วนวิธีฮิวริสติก กับวิธีของเบย์ มีความสัมพันธ์กันอย่างไม่มีนัยสำคัญทางสถิติ

ชุติมา แสงดารารัตน์ (2545) ได้เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความถนัดทางการเรียนตามความรู้สึกคุ้นเคย ความรู้สึกสนใจ และความรู้สึกพอใจในข้อสอบด้วยวิธีการตรวจแตกต่างกัน มีกลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 โรงเรียนในกรุงเทพมหานคร จำนวน 585 คน

ผลการวิจัย พบว่า จำนวนข้อสอบที่ทำหน้าต่างกันของแบบทดสอบวัดความถนัดทาง การเรียนด้านเหตุผลด้วยวิธีการตรวจแตกต่างกันสามารถบ่งชี้ข้อที่ทำหน้าที่ต่างกันจากกลุ่มอ้างอิงที่เป็นเพศ ความรู้สึกสนใจ ได้จำนวนข้อแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนกลุ่มอ้างอิงมีความรู้สึกคุ้นเคย ความรู้สึกพอใจ ได้จำนวนข้อแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สุมาลี แกล้วทนงค์ (2547) ได้ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม กับนักเรียนชั้นมัธยม ศึกษาปีที่ 1 จำนวน 1,320 คน ผลการศึกษาพบว่าข้อสอบที่ทำหน้าที่ต่างกันด้านเพศของแบบสอบสาระการเรียนรู้ภาษาไทยส่วนใหญ่มีสาเหตุมาจากความสนใจในเนื้อเรื่องและภาษาที่ใช้ในแบบสอบ ส่วนข้อสอบที่ทำหน้าที่ต่างกันด้านเพศของแบบสอบสาระการเรียนรู้สังคมศึกษาศาสนาและวัฒนธรรมนั้นมีสาเหตุมาจากเนื้อเรื่องที่สนใจและเนื้อเรื่องเกี่ยวกับวัฒนธรรม ประเพณี

อุทัยวรรณ สายพัฒนะ (2547) *ได้ศึกษาการเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST* ในเงื่อนไขกลุ่มตัวอย่างและความยาว

ของแบบทดสอบขนาดต่างๆ ผลการวิจัย พบว่า เมื่อแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST สูงกว่าวิธี GMH แต่ส่งผลต่อความผิดพลาดในการตรวจสอบในบางเงื่อนไขขนาดของกลุ่มตัวอย่าง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน

อุมาภรณ์ ภัทรวาณิชย์ และ ปัทมา อมรสิริสมบูรณ์ (2550) ได้ศึกษาการเรียนต่อระดับมัธยมปลายหรือเทียบเท่า ระหว่างปี พ.ศ. 2533 และปี พ.ศ. 2543 ชี้ให้เห็นว่า ไม่มีความแตกต่างระหว่างเพศในการเรียนต่อระดับมัธยมปลายหรือเทียบเทียบ ในปี พ.ศ. 2533 แต่ในปี พ.ศ. 2543 พบความแตกต่างระหว่างเพศ โดยเพศหญิงมีโอกาสได้เรียนต่อมากกว่าเพศชาย นอกจากนี้ ความแตกต่างระหว่างเขตเมืองและชนบท ยังเป็นปัจจัยสำคัญที่ส่งผลต่อการเรียนต่อระดับมัธยมปลายทั้งในปี พ.ศ. 2533 และ 2543 แม้ว่าความแตกต่างระหว่างเมืองและชนบทในการเรียนต่อระดับมัธยมปลายของปี พ.ศ. 2543 ได้ลดลงจากปี พ.ศ. 2533 ก็ตาม

ส่วนความไม่เท่าเทียมด้านการศึกษา: เมืองและชนบท มีความแตกต่างระหว่างเมืองและชนบท ในการจ้างงาน ความยากจน ข้อมูลจากสำนักงานสถิติแห่งชาติ (2533) ชี้ว่าประเทศไทยกำลังเพิ่มความเป็นเมืองมากขึ้น นั่นคือหนึ่งในสามของประชากรไทยอาศัยอยู่ในเขตเมือง ซึ่งเป็นการพิจารณาในระดับภาคจะยังคงพบความแตกต่างในเรื่องเมืองและชนบทอยู่

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ได้ศึกษาค่าพารามิเตอร์ข้อสอบ พารามิเตอร์ความสามารถของผู้สอบ ตรวจสอบและเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยการประยุกต์ใช้โปรแกรมโมเดลเชิงเส้นตรงระดับลัดหลั่น (HLM) และโปรแกรม BILOG – MG

ผลการวิจัย พบว่า ผลการประมาณค่าพารามิเตอร์ข้อสอบโดยโมเดล HGLM – 2L และ HGLM – 3L ด้วยสถิติ Empirical Bayesian มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG – MG ส่วนผลการประมาณค่าพารามิเตอร์ผู้สอบด้วยโมเดล HGLM – 2L

มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG – MG ส่วนโมเดล HGLM – 3L มีระดับของค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.793 ซึ่งน้อยกว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างโมเดล HGLM – 2L กับโปรแกรม BILOG – MG และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากโมเดล HGLM และจากโปรแกรม BILOG – MG ให้ผลการตรวจสอบที่เหมือนกันในทุกข้อ ทั้งนี้เป็นเพราะหลักการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจากสองวิธีมีความคล้ายคลึงกัน นั้นคือ การควบคุมให้ค่าพารามิเตอร์ความสามารถผู้สอบมีค่าคงที่ โดยโปรแกรมแต่ละโปรแกรมมีหลักการประมาณค่าดังนี้ โปรแกรม HLM ค่าความยากรายข้อถือเป็นค่าคงที่ (Fixed Effect) ซึ่งเป็นคุณสมบัติที่เหมาะสมตามหลักการสร้างข้อสอบ เพราะตรงตามคุณสมบัติความไม่ผันแปรไปตามกลุ่มผู้สอบ (Item Invariance) ส่วนค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้จากการประมาณค่าด้วยวิธีเบส์ (EB) ก็ถูกปรับให้มีค่าเฉลี่ยเท่ากับศูนย์ ส่วนเบี่ยงเบนมาตรฐานเท่ากับหนึ่ง (Normal Distribution) ดังนั้นโปรแกรมจะทดสอบอิทธิพลของตัวแปรดัมมี่เพศที่ผู้วิจัยสร้างขึ้นในสมการได้จากการทดสอบด้วยสถิติทดสอบที (*t – test*) ตามสมมติฐาน $H_0 : \gamma_{01} = 0$ หากผลการทดสอบมีนัยสำคัญก็แสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ส่วนโปรแกรม BILOG – MG จะทำการคำนวณค่าพารามิเตอร์ความสามารถของผู้สอบทั้งสองกลุ่มร่วมกัน เพื่อเป็นการปรับฐานให้อยู่บนมาตรเดียวกัน ควบคุมค่าความคลาดเคลื่อนมาตรฐานและปรับค่าเฉลี่ยค่าพารามิเตอร์ความยาก (Threshold) ของข้อสอบกลุ่มอ้างอิงให้เท่ากับศูนย์ ควบคุมค่าพารามิเตอร์การเดาให้เท่ากับศูนย์ และค่าพารามิเตอร์อำนาจจำแนกให้เท่ากันทั้งสองกลุ่ม จากนั้นจึงทำการวิเคราะห์ค่าพารามิเตอร์ความยากอีกครั้งโดยแยกวิเคราะห์ตามแต่ละกลุ่มเพื่อนำค่า – 2lnL Ratio ของแต่ละกลุ่มมาเปรียบเทียบกัน ด้วยหลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้หลักการคำนวณและวิธีการทางสถิติวิเคราะห์ที่คล้ายคลึงกัน จึงทำให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเหมือนกัน

จากการศึกษางานวิจัยที่เกี่ยวข้องที่ได้นำเสนอมาเป็นงานวิจัยประเภทดุษฎีนิพนธ์และเป็นวิทยานิพนธ์มหาบัณฑิต โดยงานวิจัยมุ่งเปรียบเทียบของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และศึกษาได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้เพศ สถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน เรื่องเป็นการหาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ จากการศึกษาสามารถประมวลความรู้ได้ว่ารูปแบบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีความเหมาะสมภายใต้เงื่อนไข และสถานการณ์ที่แตกต่างกัน ผู้ที่จะนำไปใช้ควรมีการศึกษาในรายละเอียดของเงื่อนไขให้สอดคล้องกับภาระงานของตนเอง