

การวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบ  
ตามทฤษฎีการทดสอบแบบดั้งเดิม\*  
ITEM ANALYSIS OF MULTIPLE-CHOICE QUESTIONS  
UNDER CLASSICAL TEST THEORY

ปฐิมปรัชญ์ คณินพศุศย์

Purimpratch Khaninphasut

สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

Corresponding Author E-mail: purim.orestou@gmail.com

### บทคัดย่อ

บทความวิชาการนี้มีวัตถุประสงค์เพื่อนำเสนอหลักการและแนวปฏิบัติในการวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบตามทฤษฎีการทดสอบแบบดั้งเดิม เนื่องจากคุณภาพของข้อสอบเป็นปัจจัยสำคัญต่อความตรงในการประเมินความสามารถที่แท้จริงของผู้เรียน การวิเคราะห์คุณภาพข้อสอบรายข้อจึงเป็นกระบวนการทางสถิติที่ช่วยตรวจสอบประสิทธิภาพของข้อสอบอย่างเป็นระบบ โดยครอบคลุมดัชนีสำคัญ 3 ประการ ได้แก่ 1. ค่าความยาก ซึ่งสะท้อนสัดส่วนของผู้ที่ตอบข้อสอบข้อนั้นถูก 2. ค่าอำนาจจำแนก ที่แสดงความสามารถของข้อสอบในการแยกระหว่างกลุ่มผู้สอบที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ และ 3. ประสิทธิภาพของตัวลอง ซึ่งวิเคราะห์ความสามารถของตัวลองในการดึงดูดผู้สอบกลุ่มต่ำที่ไม่มีความรู้จริงมาตอบ นอกจากนี้ บทความยังอธิบายสูตร ขั้นตอนการคำนวณ และการแปลความหมายของค่าดัชนีต่าง ๆ อย่างละเอียด รวมถึงการนำเสนอแนวทางที่ผู้สอนและนักวิจัยสามารถนำไปประยุกต์ใช้ได้จริงผ่านโปรแกรมพื้นฐาน โดยไม่จำเป็นต้องใช้ซอฟต์แวร์ทางสถิติที่ซับซ้อน การทำความเข้าใจหลักการวิเคราะห์ข้อสอบและการปรับปรุงข้อสอบให้เป็นไปตามเกณฑ์คุณภาพมาตรฐาน จะช่วยยกระดับความแม่นยำ ความเที่ยงตรง และความเป็นธรรมของการวัดผลสัมฤทธิ์ทางการเรียนให้ดียิ่งขึ้น

**คำสำคัญ:** การวิเคราะห์ข้อสอบ; ความยาก; อำนาจจำแนก; ข้อสอบแบบเลือกตอบ; ทฤษฎีการทดสอบแบบดั้งเดิม

## Abstract

This academic article aims to present principles and practical guidelines for analyzing the quality of multiple-choice questions based on Classical Test Theory. The quality of items is an important factor in the validity of assessing students' true abilities. Item analysis represents a statistical process that systematically examines the effectiveness of items through three essential indices. First, the item difficulty index reflects the proportion of examinees who answer the item correctly. Second, the item discrimination index demonstrates the ability of an item to differentiate between high-ability and low-ability examinees. Third, distractor efficiency analyzes the capability of distractors to attract low-ability examinees. Moreover, the study systematically presents the formulas, calculation procedures, and interpretive guidelines for the indices. It presents practical approaches that teachers and researchers can apply using basic software programs without requiring sophisticated statistical packages. Understanding the principles of item analysis and improving items to meet quality standards will enhance the accuracy, validity, and fairness of learning achievement.

**Keywords:** Item Analysis; Item Difficulty; Item Discrimination; Multiple-Choice Questions; Classical Test Theory

## บทนำ

แบบทดสอบแบบเลือกตอบ หรือแบบทดสอบปรนัย (Multiple Choice Questions: MCQs) เป็นเครื่องมือวัดผลสัมฤทธิ์ทางการเรียนที่ได้รับความนิยมอย่างแพร่หลายในระบบการศึกษาทั้งในและต่างประเทศ เนื่องจากสามารถวัดความรู้ความเข้าใจของผู้เรียนได้อย่างหลากหลายในเวลาจำกัด และสามารถตรวจให้คะแนนได้อย่างเป็นมาตรฐานและเที่ยงตรง (Miller et al., 2009) การใช้แบบทดสอบแบบเลือกตอบในฐานะที่เป็นเครื่องมือวัดผลการเรียนรู้จึงมีบทบาทสำคัญในการประเมินความสามารถของผู้เรียนในหลากหลายระดับการศึกษา ตั้งแต่การศึกษาขั้นพื้นฐานไปจนถึงระดับอุดมศึกษา รวมถึงการสอบวัดมาตรฐานต่าง ๆ (Haladyna et al., 2002; Rao et al., 2016)

การที่แบบทดสอบแบบเลือกตอบจะวัดความสามารถที่แท้จริงของผู้เรียนได้นั้นขึ้นอยู่กับคุณภาพของข้อสอบเป็นสำคัญ ข้อสอบที่มีคุณภาพต้องสามารถวัดสิ่งที่ต้องการวัดได้อย่างแม่นยำ มีความเที่ยงตรง และสามารถแยกความสามารถของผู้เรียนที่แตกต่าง

กันได้อย่างชัดเจน (Crocker & Algina, 1986) การสร้างข้อสอบที่ขาดคุณภาพอาจส่งผลให้การประเมินผลการเรียนรู้ไม่สะท้อนความสามารถที่แท้จริงของผู้เรียน และอาจนำไปสู่การตัดสินใจทางการศึกษาที่ไม่เหมาะสม

การตรวจสอบคุณภาพของข้อสอบแบบเลือกตอบสามารถดำเนินการได้จากการวิเคราะห์คุณภาพข้อสอบรายข้อ (Item Analysis) ซึ่งเป็นกระบวนการทางสถิติที่ช่วยให้ผู้สอนหรือผู้สร้างแบบทดสอบสามารถประเมินคุณภาพของข้อสอบแต่ละข้อได้อย่างเป็นระบบ (Ebel & Frisbie, 1991) การวิเคราะห์รายข้อนี้ประกอบด้วยดัชนีที่สำคัญ 3 ประการ ได้แก่ ค่าความยาก (Item Difficulty) ที่บอกระดับความง่ายหรือยากของข้อสอบ ค่าอำนาจจำแนก (Item Discrimination) ที่แสดงถึงความสามารถของข้อสอบในการแยกระหว่างผู้เรียนที่มีความสามารถสูงกับผู้เรียนที่มีความสามารถต่ำและค่าประสิทธิภาพของตัวลวง (Distractor Efficiency) ที่สะท้อนถึงความน่าเชื่อถือและประสิทธิผลของตัวเลือกที่ผิดในการดึงดูดผู้เรียนที่ขาดความรู้มาตอบ (Gronlund & Linn, 1990; Kehoe, 1995; ศิริชัย กาญจนวาสิ, 2556)

บทความนี้จึงมีวัตถุประสงค์เพื่ออธิบายสูตรและวิธีการคำนวณค่าความยาก ค่าอำนาจจำแนก และค่าประสิทธิภาพของตัวลวงอย่างละเอียด โดยเน้นการนำเสนอวิธีการที่ไม่ซับซ้อนและสามารถนำไปประยุกต์ใช้ได้จริง ครูผู้สอนที่สร้างแบบทดสอบเอง (Teacher-made Test) สามารถใช้วิธีการเหล่านี้ในการตรวจสอบคุณภาพของข้อสอบที่พัฒนาขึ้น (Allen & Yen, 1979) และนักวิจัยที่สร้างเครื่องมือวัดผลแบบเลือกตอบสามารถคำนวณค่าต่าง ๆ เหล่านี้ได้อย่างง่ายผ่านโปรแกรมที่มีโดยทั่วไป เช่น โปรแกรม Microsoft Excel และไม่จำเป็นต้องใช้ซอฟต์แวร์ทางสถิติที่ซับซ้อน การทำความเข้าใจและสามารถคำนวณค่าเหล่านี้ได้จะช่วยให้ผู้สอนและนักวิจัยสามารถปรับปรุงคุณภาพของข้อสอบ นำไปสู่การวัดผลสัมฤทธิ์ทางการเรียนที่มีความแม่นยำและเป็นธรรมต่อผู้เรียนมากยิ่งขึ้น

### ลักษณะของแบบทดสอบแบบเลือกตอบ

แบบทดสอบแบบเลือกตอบ (Multiple-Choice Questions) ประกอบด้วย 3 ส่วน ได้แก่ คำถาม (Stem) เป็นส่วนที่ใช้ตั้งปัญหาหรือคำถาม คำตอบที่ถูกต้อง (Keyed Response) และตัวเลือกที่ผิดตั้งแต่สองตัวเลือกขึ้นไป เรียกว่า ตัวลวง (Foils หรือ Distractors) (Crocker & Algina, 1986) เช่น แบบทดสอบแบบ 4 ตัวเลือก ประกอบด้วย คำถาม ตัวถูก 1 ตัว และตัวลวง 3 ตัว

### ตารางที่ 1 ตัวอย่างข้อสอบแบบเลือกตอบ 4 ตัวเลือก

ข้อสอบ	
0. ข้อสอบปรนัยข้อหนึ่งมีผู้เข้าสอบ 40 คน ตอบถูก 28 คน ค่าความยากของข้อสอบข้อนี้เท่ากับข้อใด	
ก. 0.30	
ข. 0.40	
ค. 0.70 (เฉลย)	
ง. 0.85	
เหตุผลของการสร้างตัวถูก/ตัวลวง	
ตัวถูก	ตัวลวง
ค่าความยาก $p = \frac{\text{จำนวนคนตอบถูก}}{\text{จำนวนคนทั้งหมด}}$ $p = \frac{28}{40} = 0.70$	ก. 0.30 เกิดจากความเข้าใจผิดว่าใช้ “จำนวนตอบผิด/ทั้งหมด = $\frac{12}{40}$ ” แทนค่าความยาก ข. 0.40 เกิดจากการคำนวณผิด เช่น หารด้วยจำนวนกลุ่มย่อย หรืออ่านข้อมูลผิด เช่น $\frac{8}{20}$ ง. 0.85 เป็นค่าที่สูงเกินจริง สะท้อนผู้เรียนที่เดาว่าข้อนี้ “ง่ายมาก” โดยไม่คำนวณ

จากตารางที่ 1 เป็นตัวอย่างของข้อสอบแบบเลือกตอบ 4 ตัวเลือก ซึ่งผู้เขียนได้อธิบายถึงที่มาของตัวถูก และตัวลวงในการสร้างข้อสอบ ครูผู้สอนหรือผู้ที่สร้างแบบทดสอบควรสร้างตัวลวงมาจากความเข้าใจที่คลาดเคลื่อนของผู้เรียน (Misconception) เพื่อเป็นสารสนเทศให้กับผู้สอนไปพัฒนาผู้เรียนที่เลือกตัวลวงเหล่านี้ต่อไป

หากครูผู้สอนนำแบบทดสอบที่สร้างขึ้นไปใช้ทดสอบกับผู้เรียน หรือผู้สร้างแบบทดสอบ/นักวิจัยนำข้อสอบไปทดลองใช้ (try out) กับกลุ่มที่ใกล้เคียงกับกลุ่มตัวอย่าง ครูผู้สอน/นักวิจัยจะได้ผลการตอบข้อสอบของผู้เรียนที่อยู่ในรูปแบบการให้คะแนน 0 กับ 1 โดยถ้าตอบผิด ได้ 0 คะแนน และตอบถูก ได้ 1 คะแนน ขั้นตอนต่อไปที่ควรดำเนินการ คือ การตรวจสอบคุณภาพของข้อสอบที่สร้างขึ้น ที่เราเรียกว่า การวิเคราะห์ข้อสอบ (Item Analysis) ดังหัวข้อที่จะกล่าวต่อไป

### การวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบ

การวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบที่นำเสนอมีพื้นฐานมาจากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) โดยการที่จะบ่งชี้ว่า ข้อสอบมีคุณภาพหรือไม่ ก็คือ การพิจารณาคุณภาพของข้อสอบรายข้อ ซึ่งพิจารณา 3 ประเด็น คือ ค่าความยาก (Item Difficulty) ค่าอำนาจจำแนก (Item Discrimination) และประสิทธิภาพตัวลวง (Distractor Efficiency)

### 1. ความยาก (Item Difficulty)

ค่าความยาก หมายถึง สัดส่วนของผู้ตอบข้อสอบข้อนั้นถูกต้องผู้สอบทั้งหมด แทนด้วยสัญลักษณ์  $p$  ความยากมีค่าตั้งแต่ 0 แต่ไม่เกิน 1 ( $0 \leq p \leq 1$ ) ถ้าสัดส่วนคนตอบข้อสอบข้อนั้นมาก แสดงว่า ข้อสอบง่าย หรือถ้าสัดส่วนคนตอบข้อสอบข้อนี้น้อย แสดงว่า ข้อสอบยาก สัดส่วนนี้เป็นไปตามหลักเหตุผล (logical) คือ ถ้าข้อสอบง่าย ผู้เรียนส่วนใหญ่ควรตอบข้อสอบข้อนั้นถูก หรือถ้าข้อสอบยาก ผู้เรียนควรตอบข้อสอบข้อนั้นได้น้อย ข้อสอบที่มีค่าความยากเหมาะสม คือ ข้อสอบที่มีค่าความยากตั้งแต่ 0.2 ถึง 0.8 ( $0.2 \leq p \leq 0.8$ ) (ศิริชัย กาญจนวาสี, 2556) การคำนวณค่าความยากสามารถคำนวณได้ด้วยสูตร

$$p_i = \frac{C_i}{N_i}$$

โดยที่  $p_i$  แทน ค่าความยากของข้อสอบข้อที่  $i$

$C_i$  แทน จำนวนผู้เรียนที่ตอบข้อสอบข้อที่  $i$  ถูก (Correct)

$N_i$  แทน จำนวนผู้เรียนที่ตอบข้อสอบข้อที่  $i$  ถูกและผิดรวมกัน หรือจำนวนผู้เรียนทั้งหมดที่ตอบข้อสอบข้อที่  $i$

**ตัวอย่างที่ 1** การคำนวณค่าความยากของข้อสอบจำนวน 3 ข้อ ผู้สอบจำนวน 10 คน

ข้อสอบข้อที่	ผู้สอบคนที่										$C_i$	$p_i$	แปลความหมาย
	1	2	3	4	5	6	7	8	9	10			
1	0	0	1	0	0	0	0	0	0	0	1	0.1	ยาก ควรปรับ
2	1	1	1	1	1	1	0	1	1	1	9	0.9	ง่าย ควรปรับ
3	1	1	1	0	1	0	1	0	1	0	6	0.6	เหมาะสม ใช้ได้

จากตัวอย่างที่ 1 การคำนวณความยากของข้อสอบข้อที่ 1

$$p_1 = \frac{C_1}{N_1} = \frac{1}{10} = 0.1$$

สำหรับข้ออื่น ๆ ก็คำนวณเหมือนกัน โดยคำนวณทีละข้อ แต่สิ่งสำคัญมิใช่การคำนวณได้ตัวเลขมาเท่านั้น ครูผู้สอนหรือผู้วิจัยต้องแปลความหมายค่าความยากได้ว่า ข้อสอบข้อใดมีค่าความยากเหมาะสม นำไปใช้ได้ ไม่ต้องปรับ หรือข้อสอบข้อใดยากเกินไป หรือง่ายเกินไป ควรปรับดังนี้ เช่น ข้อ 1 ข้อสอบยากเกินไป ควรปรับ เพราะมีคนตอบถูกเพียง 1 คนจาก 10 คน (10%) ทำให้อาจไม่สามารถวัดความรู้ที่แท้จริงได้ ผู้สอบส่วนใหญ่อาจจะเดา หรือเนื้อหาอาจจะมีความสับสนระดับการเรียนรู้ของผู้เรียน ดังนั้นควรนำข้อนี้ไปปรับปรุงแก้ไขเนื้อหา คำถาม หรือตรวจสอบว่าเฉลยผิดหรือไม่

## 2. อำนาจจำแนก (Item Discrimination)

ค่าอำนาจจำแนก หมายถึง ข้อสอบสามารถแยกผู้สอบที่มีความสามารถสูง (กลุ่มสูง/คะแนนสูง) ออกจากผู้สอบที่มีความสามารถต่ำ (กลุ่มต่ำหรือคะแนนต่ำ) แทนด้วยสัญลักษณ์  $D$  (Crocker & Algina, 1986) หรือ  $r$  (ศิริชัย กาญจนวาสี, 2556) ในที่นี้ผู้เขียนขอใช้สัญลักษณ์  $r$  ค่าอำนาจจำแนกมีค่าตั้งแต่  $-1$  ถึง  $1$  ( $-1 \leq r \leq 1$ ) ข้อสอบที่มีค่าอำนาจจำแนกเหมาะสม คือ ข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่  $0.2$  ขึ้นไป ( $r \geq 0.2$ ) (ศิริชัย กาญจนวาสี, 2556)

ในการคำนวณค่าอำนาจจำแนก จะต้องนำคะแนนรวมของผู้สอบแต่ละคนมาเรียงลำดับจากมากไปหาน้อย แล้วใช้เกณฑ์ในการแบ่งผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มสูง (คะแนนสูง) กับกลุ่มต่ำ (คะแนนน้อย) ซึ่งเกณฑ์ที่นักวิจัยส่วนใหญ่ใช้ในการแบ่ง คือ กลุ่มสูง 27% และกลุ่มต่ำ 27% กรณีที่คะแนนของผู้สอบมีการแจกแจงแบบปกติ (Normal Distribution) (Kelly, 1939 อ้างถึงใน Crocker & Algina, 1986; Krishnan, 2013) แต่ถ้าจำนวนผู้สอบมีขนาดเล็ก อาจใช้เกณฑ์แบ่งครึ่ง คือ กลุ่มสูง 50% และกลุ่มต่ำ 50% (Crocker & Algina, 1986) การคำนวณอำนาจจำแนกสามารถคำนวณได้ด้วยสูตร

$$r_i = \frac{U_i}{n_{iU}} - \frac{L_i}{n_{iL}}$$

โดยที่  $r_i$  แทน ค่าอำนาจจำแนกข้อที่  $i$   
 $U_i$  แทน จำนวนผู้สอบที่อยู่ในกลุ่มสูงตอบข้อสอบข้อที่  $i$  ถูก  
 $L_i$  แทน จำนวนผู้สอบที่อยู่ในกลุ่มต่ำตอบข้อสอบข้อที่  $i$  ถูก  
 $n_{iU}$  แทน จำนวนผู้สอบที่อยู่ในกลุ่มสูงของข้อสอบข้อที่  $i$   
 $n_{iL}$  แทน จำนวนผู้สอบที่อยู่ในกลุ่มต่ำของข้อสอบข้อที่  $i$

ถ้า  $n_{iU} = n_{iL} = n_i$  สามารถเขียนสูตรอยู่ในรูป

$$r_i = \frac{U_i - L_i}{n_i}$$

**ตัวอย่างที่ 2** การคำนวณค่าอำนาจจำแนกข้อสอบ 4 ข้อ ผู้สอบทั้งหมด 100 คน ใช้เกณฑ์กลุ่มสูง 27% และกลุ่มต่ำ 27% จะได้

กลุ่มสูง ( $U$ : Upper) จำนวน 27 คน (นักเรียนที่ได้คะแนนรวมสูงที่สุด 27 อันดับแรก)

กลุ่มต่ำ ( $L$ : Lower) จำนวน 27 คน (นักเรียนที่ได้คะแนนรวมน้อยที่สุด 27 อันดับสุดท้าย)

ข้อที่	กลุ่มสูงตอบถูก ( $U_i$ )	กลุ่มต่ำตอบถูก ( $L_i$ )	การคำนวณ $r_i$	ค่า $r_i$	แปลความหมาย
1	22	5	$\frac{22 - 5}{27}$	0.63	เหมาะสม ใช้ได้
2	10	18	$\frac{10 - 18}{27}$	-0.3	ติดลบ ควรปรับปรุง
3	12	12	$\frac{12 - 12}{27}$	0	จำแนกไม่ได้ ควรปรับปรุง
4	15	12	$\frac{15 - 12}{27}$	0.11	จำแนกได้ต่ำ ควรปรับปรุง

จากตัวอย่างที่ 2 เป็นการคำนวณและแปลผลค่าอำนาจจำแนกของข้อสอบแต่ละข้อ เมื่อพิจารณาค่า  $r_i$  จะเห็นว่า มีค่าอำนาจจำแนกที่เป็นบวก (มากกว่า 0.2/ต่ำกว่า 0.2) ติดลบ และศูนย์ ซึ่งทั้งสามกรณีนี้มีความหมายที่ต่างกัน ดังนี้

ข้อที่ 1 ค่า  $r = 0.63$  ซึ่งมากกว่า 0.2 แสดงว่า อยู่ในเกณฑ์ที่กำหนด จำแนกผู้สอบกลุ่มสูงกับกลุ่มต่ำออกจากกันได้ชัดเจน สังเกตข้อมูลตารางที่ 2 ( $U = 22, L = 5$ ) จึงสรุปว่า ข้อนี้ใช้ได้

ข้อที่ 2 ค่า  $r = -0.3$  อำนาจจำแนกติดลบ หมายความว่า กลุ่มต่ำตอบถูกมากกว่ากลุ่มเก่ง หรือ คนอ่อนตอบถูกแต่คนเก่งตอบผิด สังเกตข้อมูลตารางที่ 2 ( $U = 10, L = 18$ ) จึงสรุปว่า ข้อนี้ใช้ไม่ได้ ควรปรับปรุงหรือตัดทิ้งในกรณีที่ข้อสอบมีจำนวนมาก ทั้งนี้ การที่คนเก่งตอบผิดอาจเกิดจากเฉลยผิด คำถามกำกวมที่ทำให้คนเก่งคิดลึกเกินไป หรือตัวลวงดึงดูดคนเก่งมากกว่าคนอ่อน ดังนั้น ควรตรวจสอบเฉลย และเนื้อหาใหม่อีกครั้ง

ข้อที่ 3 ค่า  $r = 0.00$  อำนาจจำแนกเท่ากับ 0 หมายถึง ข้อสอบจำแนกคนสูงออกจากกลุ่มต่ำไม่ได้เลย หรือกลุ่มสูงและกลุ่มต่ำมีความสามารถในข้อนี้เท่ากัน (ตอบถูกเท่ากัน) สังเกตข้อมูลตารางที่ 2 ( $U = 12, L = 12$ ) จึงสรุปว่า ข้อนี้ใช้ไม่ได้ ควรปรับปรุง

ข้อที่ 4 ค่า  $r = 0.11$  อำนาจจำแนกต่ำกว่า 0.2 แต่เป็นบวก ข้อสอบจำแนกได้น้อย สังเกตข้อมูลตารางที่ 2 ( $U = 15, L = 12$ ) แม้ว่าคนเก่งจะตอบถูกมากกว่าคนอ่อนเล็กน้อย แต่ก็ยังไม่ผ่านเกณฑ์ที่กำหนด จึงสรุปว่า ข้อนี้ใช้ไม่ได้ ควรปรับปรุง โดยอาจปรับปรุงตัวเลือกให้มีประสิทธิภาพมากขึ้นเพื่อให้กลุ่มอ่อนตอบถูกมากขึ้น

### 3. ประสิทธิภาพตัวลวง (Distractor Efficiency)

ประสิทธิภาพของตัวลวง คือ ความสามารถของตัวเลือกที่ไม่ใช่คำตอบถูก (Distractors/Foils) ในการดึงดูดผู้สอบกลุ่มต่ำ (กลุ่มอ่อน) ให้เลือกตอบ ในขณะที่ผู้สอบกลุ่มสูง (กลุ่มเก่ง) สามารถแยกแยะได้ว่าตัวเลือกนั้นไม่ถูกต้อง ส่วนใหญ่แทนด้วยสัญลักษณ์ DE การคำนวณประสิทธิภาพตัวลวงสามารถคำนวณได้ด้วยสูตร

สัดส่วนของผู้เลือกตัวลง (ศิริชัย กาญจนวาสี, 2556)

$$p_w = \frac{W_U + W_L}{N_U + N_L}$$

หรือ กรณีที่  $N_U = N_L$

$$p_w = \frac{p_{wU} + p_{wL}}{2}$$

อำนาจจำแนกของตัวลง (ศิริชัย กาญจนวาสี, 2556)

สูตรนี้ให้สังเกตว่า เราจะนำจำนวนผู้สอบในกลุ่มต่ำเป็นตัวตั้ง แล้วนำไปลบออกจากจำนวนผู้สอบกลุ่มสูง ค่าอำนาจจำแนกของตัวลงจึงควรมีค่าเป็นบวก

$$r_w = \frac{W_L - W_U}{N_L \text{ หรือ } N_U}$$

หรือ

$$r_w = p_{wL} - p_{wU}$$

โดยที่	$p_w$	แทน สัดส่วนของผู้เลือกตัวลงนั้น
	$r_w$	แทน อำนาจจำแนกของตัวลงนั้น
	$W_U$	แทน จำนวนผู้สอบที่อยู่ในกลุ่มสูงที่เลือกตัวลงนั้น
	$W_L$	แทน จำนวนผู้สอบที่อยู่ในกลุ่มต่ำที่เลือกตัวลงนั้น
	$N_U$	แทน จำนวนผู้สอบที่อยู่ในกลุ่มสูง
	$N_L$	แทน จำนวนผู้สอบที่อยู่ในกลุ่มต่ำ
	$p_{wL} = \frac{W_L}{N_L}$	และ $p_{wU} = \frac{W_U}{N_U}$

หากพิจารณาสูตรการคำนวณ ค่าสัดส่วนของตัวลงมีค่าตั้งแต่ 0 ถึง 1 ( $0 \leq p_w \leq 1$ ) และค่าอำนาจจำแนกของตัวลงมีค่าตั้งแต่ -1 ถึง 1 ( $-1 \leq r_w \leq 1$ ) สำหรับการที่จะตัดสินว่า ตัวลงนั้นทำหน้าที่ได้ (functioning distractor: FD) ตัวลงนั้นจะมีค่าสัดส่วนของผู้เลือกตัวลงตั้งแต่ 0.05 ขึ้นไป ( $p_w \geq 0.05$ ) (ศิริชัย กาญจนวาสี, 2556) และมีค่าอำนาจจำแนกของตัวลงตั้งแต่ 0.05 ขึ้นไปเช่นกัน ( $r_w \geq 0.05$ ) (ศิริชัย กาญจนวาสี, 2556)

**ตัวอย่างที่ 3** การคำนวณประสิทธิภาพตัวลง ผู้สอบทั้งหมด 100 คน ใช้เกณฑ์กลุ่มสูง 27% และกลุ่มต่ำ 27% จะได้

กลุ่มสูง ( $U$ : Upper) จำนวน 27 คน (นักเรียนที่ได้คะแนนรวมสูงที่สุด 27 อันดับแรก)

กลุ่มต่ำ ( $L$ : Lower) จำนวน 27 คน (นักเรียนที่ได้คะแนนรวมน้อยที่สุด 27 อันดับสุดท้าย)

ตัวเลือก	กลุ่มสูง ( $N_U$ )	กลุ่มต่ำ ( $N_L$ )	รวม $N_U + N_L$	$p$	$r$	แปลความหมาย
ก. (ถูก)	20	5	25	0.46	0.56	ข้อสอบใช้ได้
ข. (ลวง)	2	8	10	0.19	0.22	ตัวลวงใช้ได้
ค. (ลวง)	4	4	8	0.15	0	ตัวลวงใช้ไม่ได้เพราะ $r_w = 0$
ง. (ลวง)	1	10	11	0.20	0.33	ตัวลวงใช้ได้

จากตัวอย่างที่ 3 เมื่อพิจารณาเฉพาะตัวลวง ข, ค, ง การคำนวณค่า  $p_w$  และ  $r_w$

คือตัวลวง ข  $p_w = \frac{2+8}{27+27} = \frac{10}{54} = 0.19$  ผ่านเกณฑ์ เพราะมากกว่า 0.05

$r_w = \frac{8-2}{27} = \frac{6}{27} = 0.22$  ผ่านเกณฑ์ เพราะมีค่าเป็นบวก และมากกว่า 0.05

ดังนั้น ตัวลวง ข เป็นตัวลวงที่มีประสิทธิภาพ

ตัวลวง ค  $p_w = \frac{4+4}{27+27} = \frac{8}{54} = 0.15$  ผ่านเกณฑ์ เพราะมากกว่า 0.05

$r_w = \frac{4-4}{27} = \frac{0}{27} = 0$  ไม่ผ่านเกณฑ์ เพราะมีค่าเป็น 0 ซึ่งน้อยกว่า 0.05

ดังนั้น ตัวลวง ค ควรปรับปรุง เพราะเป็นตัวลวงที่ไม่มีอำนาจจำแนก คือลวงผู้สอบทั้งสองกลุ่มได้เท่ากัน

ตัวลวง ง  $p_w = \frac{1+10}{27+27} = \frac{11}{54} = 0.20$  ผ่านเกณฑ์ เพราะมากกว่า 0.05

$r_w = \frac{10-1}{27} = \frac{9}{27} = 0.33$  ผ่านเกณฑ์ เพราะมีค่าเป็นบวก และมากกว่า 0.05

ดังนั้น ตัวลวง ง เป็นตัวลวงที่มีประสิทธิภาพ โดยเป็นตัวลวงที่ทำหน้าที่ดีที่สุดในข้อสอบข้อนี้

สำหรับตัวเลือก ก ซึ่งเป็นตัวถูก เราจะนำค่าความยาก และค่าอำนาจจำแนกมาพิจารณาร่วมกัน เพื่อตัดสินคุณภาพของข้อสอบข้อนี้ว่า ข้อสอบใช้ได้ ควรปรับปรุง หรือควรตัดทิ้ง ซึ่งทั้งสองค่านี้ต้องผ่านเกณฑ์ทั้งคู่จึงจะถือว่า ข้อสอบนำไปใช้ได้ ดังนั้น จากตัวเลือก ก ค่าความยากเท่ากับ 0.46 และค่าอำนาจจำแนกเท่ากับ 0.56 จึงสรุปได้ว่า ข้อสอบข้อนี้นำไปใช้ได้

หากมองอีกมุมหนึ่ง ถ้าเลือกใช้แนวคิดของ Functioning Distractor (FD) ซึ่งเป็นที่นิยมใช้รายงานในบทความวิชาการระดับนานาชาติ เกณฑ์การพิจารณาจะมุ่งเน้นไปที่ การทำหน้าที่ของตัวลวงในเชิงสถิติ หากผู้เรียนเลือกตัวลวงตั้งแต่ร้อยละ 5 ขึ้นไป จะเรียกว่า ตัวลวงที่ทำหน้าที่ได้ (Functioning Distractor: FD) ส่วนผู้เรียนเลือกตัวลวง น้อยกว่าร้อยละ 5 จะเรียกว่า ตัวลวงที่ไม่ทำหน้าที่ (Non-Functioning Distractor: NFD) (Reziggalla et al., 2024; Tarrant, Ware et al., 2009) ตัวอย่างการคำนวณเป็น ดังนี้

สมมติข้อสอบแบบเลือกตอบ 4 ตัวเลือก มีผู้สอบ 40 คน สามารถจำแนกจำนวนผู้ตอบตัวเลือกทั้ง 4 ตัวเลือกได้ ดังนี้

ตัวเลือก	จำนวนผู้เลือก	ร้อยละ	สถานะ
ก. ตัวลวง	10	25%	FD
ข. ตัวถูก	22	55%	-
ค. ตัวลวง	6	15%	FD
ง. ตัวลวง	2	5%	FD

หมายเหตุ FD (functioning distractor) หมายถึง ตัวลวงทำหน้าที่ได้

จากตารางข้างต้น การคำนวณร้อยละของการทำหน้าที่ของตัวลวง ก, ค, ง คือ

ตัวลวง ก ร้อยละ =  $\frac{10}{40} \times 100 = 25\%$  มากกว่าร้อยละ 5 ตัวลวงทำหน้าที่ได้ (FD)

ตัวลวง ค ร้อยละ =  $\frac{6}{40} \times 100 = 15\%$  มากกว่าร้อยละ 5 ตัวลวงทำหน้าที่ได้ (FD)

ตัวลวง ง ร้อยละ =  $\frac{2}{40} \times 100 = 5\%$  เท่ากับร้อยละ 5 ตัวลวงทำหน้าที่ได้ (FD)

ดังนั้น ตัวลวงทั้งหมด 3 ตัว ทำหน้าที่ของตัวลวงได้ทั้งหมด

#### การคำนวณประสิทธิภาพตัวลวง (Distractor Efficiency: DE)

ค่า DE อยู่ในช่วงตั้งแต่ 0-100% ซึ่งใช้สะท้อนคุณภาพของตัวลวงในข้อสอบแบบเลือกตอบ (Gierl & Bulut, 2017)

$$DE = \frac{\text{จำนวนตัวลวงที่ทำหน้าที่ได้ (FD)}}{\text{จำนวนตัวลวงทั้งหมด}} \times 100$$

จากตัวอย่างข้างต้น จำนวนตัวลวงที่ทำหน้าที่ได้ = 3 ตัว (FD = 3)

แทนค่าในสูตร  $DE = \frac{3}{3} \times 100 = 100\%$

ดังนั้น ข้อสอบข้อนี้มีประสิทธิภาพตัวลวงสูงมาก (DE = 100%)

จากการพิจารณาประสิทธิภาพตัวลวงของทั้งสองวิธีต่างก็มุ่งแสดงให้เห็นว่าตัวลวงที่สร้างขึ้นสามารถทำหน้าที่ได้ดีหรือไม่ คือ ลวงคนอ่อนมาตอบมากกว่าคนเก่ง เพราะแนวคิดตัวลวงคือ ตัวเลือกที่สร้างมาจากความเข้าใจที่คลาดเคลื่อนในเนื้อหาของผู้เรียน หากผู้เรียนที่ได้คะแนนต่ำ เลือกตัวลวงเหล่านี้ก็แสดงว่า ยังไม่บรรลุวัตถุประสงค์การเรียนรู้ในเรื่องที่ออกข้อสอบ สำหรับการเลือกใช้แต่ละวิธี หากต้องการเน้นคัดกรองตัวลวงเบื้องต้นเพื่อตัดตัวลวงที่ไม่มีใครเลือกออก ก็สามารถใช้เกณฑ์ Functioning Distractor (FD) ที่พิจารณาร้อยละการเลือกตั้งแต่ 5% เป็นหลัก ซึ่งจะช่วยลดโอกาสที่ผู้สอบจะเดาถูกโดยการตัดตัวเลือกที่เด่นชัดว่าไม่ใช่คำตอบออกไป แต่ถ้าหากต้องการเน้นการจำแนกผู้สอบ ควรใช้สูตรการคำนวณค่าอำนาจจำแนกตัวลวง (r) ตามแนวคิดดัชนีกลุ่มสูง-กลุ่มต่ำ วิธีนี้จะช่วยให้เห็นว่าตัวลวงนั้นทำหน้าที่หลอกหรือลวงผู้ที่มีความรู้ร้น้อย

ได้จริงหรือไม่ หรือลงคนเก่งมาตอบมากกว่าคนอ่อน ซึ่งเป็นหัวใจสำคัญของการสร้างข้อสอบที่ต้องการความแม่นยำในการแบ่งระดับความสามารถ

แนวทางปฏิบัติในการวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบ ดำเนินการ ดังนี้

1. เรียงลำดับคะแนนรวมจากน้อยไปหามาก
2. คำนวณจำนวนนักเรียนในกลุ่มสูงและกลุ่มต่ำ โดยเลือกที่จะใช้เกณฑ์ใด เช่น กลุ่มสูง 27% และกลุ่มต่ำ 27% หรือแบ่งครึ่งกรณีจำนวนผู้สอบทั้งหมดมีจำนวนน้อย
3. แบ่งกระดาษคำตอบของผู้เรียนโดยแบ่งเป็นกลุ่มสูง และกลุ่มต่ำ ตามจำนวนที่คำนวณได้ในข้อ 2
4. จัดทำตารางแสดงจำนวนนักเรียนในกลุ่มสูงและกลุ่มต่ำที่ตอบตัวเลือกต่าง ๆ
5. คำนวณค่าความยากของข้อสอบแต่ละข้อ (คำนวณตัวถูก/เฉลย)
6. คำนวณค่าอำนาจจำแนกของข้อสอบแต่ละข้อ (คำนวณตัวถูก/เฉลย)
7. คำนวณประสิทธิภาพของตัวลองของข้อสอบแต่ละข้อ

## สรุป

การวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบ จะพิจารณาใน 3 ประเด็น คือ ค่าความยาก ค่าอำนาจจำแนก และประสิทธิภาพตัวลอง โดยเริ่มจากการพิจารณาว่าข้อสอบข้อนั้นใช้ได้หรือไม่ ให้พิจารณาจากค่าความยากและค่าอำนาจจำแนกของตัวถูก ซึ่งจะบอกถึงคุณภาพข้อสอบข้อนั้น ต่อมาให้พิจารณาประสิทธิภาพตัวลองว่า มีผู้สอบเลือกตอบหรือไม่ ถ้าไม่มีผู้สอบเลือกเลย แสดงว่า ตัวลองนั้นไม่ทำหน้าที่ลงคนอ่อนหรือมีความรู้่น้อยมาตอบ ต้องปรับปรุงตัวลอง เช่น อาจพิจารณาว่า ตัวลองนั้นเด่นชัดจนเกินไป ทำให้ผู้สอบสามารถตัดทิ้งได้หรือไม่ ถ้ามีตัวลองลักษณะนี้ควรปรับปรุงหรือตัดทิ้ง

## เอกสารอ้างอิง

- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, CA.: Brooks/Cole.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart and Winston.
- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

- Gierl, M. J. & Bulut, O. (2017). Using distractor analysis to evaluate item quality. In M. J. Gierl & O. Bulut (Eds.), *Handbook of diagnostic classification models*. (pp. 81–108). Cham: Springer.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York, NY: Macmillan.
- Haladyna, T. M. et al. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-333.
- Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, 4(10), 1-3.
- Kelly, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology*, 30, 17-24.
- Krishnan, D. R. (2013). *Statistical estimation techniques requiring nearly normal sampling distributions*. Retrieved February 1, 2026, from [https://www.pure.ed.ac.uk/ws/files/29266196/frp0477\\_krishnan.pdf](https://www.pure.ed.ac.uk/ws/files/29266196/frp0477_krishnan.pdf)
- Miller, M. D. et al. (2009). *Measurement and assessment in teaching* (10th ed.). Upper Saddle River, NJ: Pearson.
- Rao, C. et al. (2016). Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, 2(4), 201-204.
- Rezigalla, A. A. et al. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Med Educ*, 24(445), 1-7.
- Tarrant, M. et al. (2009) An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Med Educ*, 9(40), 1-8.