คู่มือผู้ใช้งานระบบวิเคราะห์คุณภาพข้อสอบ

STATNRRU ITEM ANALYSIS PRO

คู่มือรุ่น 20261020 selection13

ปรับปรุงวันที่ 20 ตุลาคม 2569

เอกสารนี้อธิบายการเตรียมข้อมูล การวิเคราะห์อิงกลุ่มและอิงเกณฑ์ การคัดเลือกข้อสอบด้วยตนเองและตาม Blueprint ตลอดจนการจัดทำรายงาน โดยใช้ชื่อเมนูและพฤติกรรมของเว็บไซต์รุ่นที่ระบุ
ผู้ใช้ควรอ่านบทการนำเข้าและการกำหนดขอบเขตก่อนวิเคราะห์ข้อมูลจริง

การนำเข้า การวิเคราะห์ การคัดเลือก และการจัดทำรายงาน

สารบัญ

01 ภาพรวมและขอบเขตการใช้งาน 3

02 เริ่มต้นใช้งานและลำดับงาน 4

03 เตรียมไฟล์และเลือกรูปแบบตาราง 5

04 ตัวอย่างการนำเข้าและการตรวจจำนวน 6

05 โหมดคะแนน คำตอบดิบ และไฟล์เฉลย 7

06 ตรวจตัวอย่างข้อมูลและแก้ไขเบื้องต้น 8

07 บันทึกและเรียกคืน JSON 9

08 กำหนดมิติย่อยและระดับ Bloom 10

09 แถบวิเคราะห์และความหมายของขอบเขต 11

10 วิเคราะห์แบบอิงกลุ่ม 12

11 วิเคราะห์แบบอิงเกณฑ์ 13

12 อ่านผลสรุป ตาราง และกราฟ 14

13 เลือกอัตโนมัติ เลือกเอง และคำนวณใหม่ 15

14 จัดชุดตาม Blueprint 16

15 วิเคราะห์ตัวเลือก ตัวลวง และ Trace Line 17

16 จัดทำรายงานมาตรฐานและส่งออก 18

17 ออกแบบ Printout ให้เหมาะกับงาน 19

18 สูตรและฐานคำนวณที่ควรทราบ 20

19 เกณฑ์แปลผลและการใช้ในงานวิจัย 21

20 ชุดตัวอย่างสำหรับการสอน 22

21 การแจ้งเตือนและการแก้ปัญหา 23

22 ติดตั้ง สำรองงาน และตรวจรับ 24

23 เอกสารอ้างอิง 25

01 ภาพรวมและขอบเขตการใช้งาน

ระบบใช้สถิติตามทฤษฎีการทดสอบแบบดั้งเดิมเพื่อช่วยตรวจทานข้อสอบ ไม่ใช่เครื่องมือรับรองความตรง ความเป็นมิติเดียว ความเป็นธรรม หรือความเหมาะสมของคะแนนเพื่อการตัดสินใจทุกประเภท

รองรับคะแนนรายข้อแบบ 0/1 คะแนนหลายระดับ และคำตอบรายตัวเลือกพร้อมเฉลย
ข้อมูลคำตอบถูกให้คะแนนแบบตรงเฉลยเท่ากับ 1 และไม่ตรงเฉลยเท่ากับ 0
ไม่รองรับการให้เครดิตบางส่วนจากคำตอบตัวเลือกหลายคำตอบโดยอัตโนมัติ

การอ่านไฟล์และคำนวณดำเนินการในเบราว์เซอร์ เว็บไซต์รุ่นนี้ไม่ใช้ Google Sheets
และไม่มีฐานข้อมูลสำหรับเก็บประวัติผู้ใช้ การปิดแท็บหรือรีเฟรชอาจทำให้ข้อมูลที่ยังไม่บันทึกสูญหาย

เว็บไซต์โหลดไลบรารีและบางฟอนต์จาก CDN จึงควรมีอินเทอร์เน็ต
การประมวลผลในเครื่องไม่ได้หมายความว่าเป็นระบบออฟไลน์สมบูรณ์หรือผ่านการตรวจรับด้านความมั่นคงปลอดภัยทุกมิติ

ควรใช้รหัสแทนชื่อจริงเมื่อทำได้ ไฟล์รายงาน ข้อมูลดิบ และ JSON อาจมีข้อมูลรายวิชา ชื่อผู้รับผิดชอบ หรือข้อมูลผู้สอบ ผู้เผยแพร่ต้องตรวจสิทธิ์เข้าถึงและองค์ประกอบที่เลือกส่งออก

ข้อจำกัดสำคัญคือไม่มี IRT การเทียบคะแนนต่างฉบับ การประมาณช่วงความเชื่อมั่นของดัชนี การวิเคราะห์ความสม่ำเสมอของการตัดสินผ่าน หรือ conditional SEM บริเวณจุดตัด ห้ามอ่าน Livingston k^2 เป็นร้อยละความถูกต้องในการจำแนก

02 เริ่มต้นใช้งานและลำดับงาน

1. เปิดเว็บไซต์ กรอกโรงเรียนหรือหน่วยงาน รายวิชา ผู้รับผิดชอบ และปีการศึกษาตามต้องการ
ช่องเหล่านี้ใช้ประกอบรายงาน ไม่ใช่ข้อมูลคะแนน

2. นำเข้าไฟล์ข้อมูล หากมี JSON เดิม สามารถนำเข้า JSON ก่อนข้อมูล หรือโหลดข้อมูลแล้วนำเข้า JSON ตามได้

3. เลือกแนวแถวหรือแนวคอลัมน์ ตำแหน่งรหัสผู้สอบ และตำแหน่งหัวข้อให้ตรงกับไฟล์
ตรวจจำนวนผู้สอบและจำนวนข้อที่ระบบแสดง

4. ตรวจคะแนนเต็มหรือเฉลย กำหนดมิติย่อยและ Bloom แล้วตรวจข้อมูลตัวอย่าง

5. กำหนดวิธีวิเคราะห์และขอบเขต ตรวจจุดตัดเมื่อใช้อิงเกณฑ์ แล้วกดวิเคราะห์และยืนยัน

6. ตรวจผล คัดเลือกข้อ หรือจัดชุดตาม Blueprint หากต้องการผลของชุดใหม่ให้สั่งวิเคราะห์ชุดที่เลือกอีกครั้ง

7. เปิดหน้ารายงาน HTML ตรวจองค์ประกอบและรูปแบบ แล้วจึงบันทึก HTML ส่งออก Word หรือพิมพ์เป็น PDF

แถบขั้นตอนด้านบนเป็นทางลัด ไม่จำเป็นต้องทำทุกขั้นใหม่เมื่อปรับเพียงเกณฑ์ ใช้ “ปรับวิธี/เกณฑ์” ในหน้าผลได้โดยตรง ส่วน “แก้ข้อมูล/โครงสร้าง” ใช้เมื่อต้องแก้เฉลย คะแนนเต็ม มิติ หรือข้อมูลต้นทาง

03 เตรียมไฟล์และเลือกรูปแบบตาราง

รองรับ CSV ไฟล์ข้อความคั่นด้วย comma semicolon หรือ tab และ Excel .xlsx/.xls โดยอ่านชีตแรกเท่านั้น หากมีหลายชีตให้จัดข้อมูลที่จะวิเคราะห์ไว้ในชีตแรก ตรวจค่าที่อ่านได้จากสูตร Excel และหลีกเลี่ยงเซลล์ผสานหรือแถวคำอธิบายก่อนตาราง

คะแนนต้องเป็นตัวเลขไม่ติดลบและไม่เกินคะแนนเต็ม ใช้จุดเป็นทศนิยม ไม่ใส่เครื่องหมายร้อยละหรือหน่วยในช่องคะแนน ระบบไม่เติมค่าขาดหายหรือแทนช่องว่างด้วยศูนย์ให้เองในโหมดคะแนน

การวางผู้สอบ	รหัสผู้สอบ	หัวข้อหรือเลขข้อ
เป็นแถว	ไม่มีคอลัมน์รหัส	ไม่มีหัวข้อ
เป็นแถว	มีคอลัมน์แรกเป็นรหัส	ไม่มีหัวข้อ
เป็นแถว	ไม่มีคอลัมน์รหัส	แถวแรกเป็นชื่อข้อ
เป็นแถว	มีคอลัมน์แรกเป็นรหัส	แถวแรกเป็นชื่อข้อ
เป็นคอลัมน์	ไม่มีแถวรหัส	ไม่มีหัวข้อ
เป็นคอลัมน์	มีแถวแรกเป็นรหัส	ไม่มีหัวข้อ
เป็นคอลัมน์	ไม่มีแถวรหัส	คอลัมน์แรกเป็นชื่อข้อ
เป็นคอลัมน์	มีแถวแรกเป็นรหัส	คอลัมน์แรกเป็นชื่อข้อ

ตัวเลือกเรื่องรหัสและหัวข้อเป็นคนละเรื่อง ต้องเลือกแยกกัน รุ่นนี้ไม่เดาหรือลบหัวตารางอัตโนมัติ การเลือก “ไม่มีหัวข้อ” หมายถึงเก็บแถวหรือคอลัมน์นั้นเป็นข้อมูลจริง

04 ตัวอย่างการนำเข้าและการตรวจจำนวน

ตัวอย่างผู้สอบเป็นแถว มีคอลัมน์รหัสและแถวหัวข้อ ให้เลือก “ผู้สอบอยู่เป็นแถว” เปิดตัวเลือกรหัส และเลือก “แถวแรกเป็นชื่อข้อ/เลขข้อ”

ID,Q1,Q2

S01,1,0

S02,0,1

ผลที่ต้องได้คือ 2 คน × 2 ข้อ รหัส S01 และ S02 ต้องยังอยู่ ระบบแสดงว่าต้นทางมี 3 แถว × 3 คอลัมน์และแยกหัวข้อ 1 แถว ไม่ใช่ลบผู้สอบคนแรก

ตัวอย่างผู้สอบเป็นคอลัมน์ มีทั้งแถวรหัสและคอลัมน์เลขข้อ ให้เปิดตัวเลือกรหัสและเลือก “คอลัมน์แรกเป็นชื่อข้อ/เลขข้อ”

Item,S01,S02

Q1,1,0

Q2,0,1

หากตารางแนวคอลัมน์มีแถว S01,S02 แล้วตามด้วยคะแนน โดยไม่มีคอลัมน์ Item ให้เปิดเพียง “มีแถวแรกเป็นรหัส/ชื่อผู้สอบ” และเลือก “ไม่มีหัวข้อหรือเลขข้อ” ไม่ต้องสมมติว่ามีคอลัมน์เลขข้อเพิ่ม

ตรวจข้อความสรุปการนำเข้า จำนวนข้อ แถวตัวอย่างแรก และแถวท้ายเสมอ
การเลือกแนวตารางผิดอาจได้ตารางที่ยังคำนวณได้แต่ความหมายผิด
ระบบตรวจจำนวนคอลัมน์ไม่เท่ากันและแจ้งแถวที่ผิด โดยไม่ตัดคอลัมน์ส่วนเกินทิ้งเงียบ ๆ

สำหรับรายงานที่สร้างจากรุ่นก่อนแก้ปัญหาหัวตาราง ควรนำเข้าไฟล์ต้นฉบับใหม่ ตรวจจำนวนคน แล้ววิเคราะห์และจัดทำรายงานใหม่ ข้อมูลที่เคยถูกตัดออกไม่สามารถกู้คืนได้จากรายงานหรือ JSON การตั้งค่าเพียงอย่างเดียว

05 โหมดคะแนน คำตอบดิบ และไฟล์เฉลย

“เป็นไฟล์คำตอบดิบ” ใช้เมื่อแต่ละช่องเป็นคำตอบ เช่น A B C D หรือ ก ข ค ง ถ้าใช้ตัวเลขแทนตัวเลือก เช่น 1 2 3 4 ต้องตรวจและเปิดโหมดคำตอบดิบเอง ระบบอาจอ่านเป็นคะแนนเพราะทุกช่องเป็นตัวเลข

ในโหมดคะแนนไม่ต้องนำเข้าเฉลย ให้ระบุคะแนนเต็มจริงของแต่ละข้อ
อย่าใช้คะแนนสูงสุดที่พบในกลุ่มตัวอย่างแทนคะแนนเต็มตามแบบทดสอบโดยไม่ตรวจสอบ

ในโหมดคำตอบดิบ เลือกนำเข้าไฟล์เฉลยแล้วตรวจตารางรายข้อ รองรับรูปแบบแนวนอนดังนี้

Key,A,B,C,D

ใช้ A,B,C,D โดยไม่มีคำว่า Key ก็ได้ หรือใช้แนวตั้งที่จับคู่เลขข้อกับคำตอบดังนี้

Item,Key

1,A

2,B

3,C

4,D

อีกแบบคือคอลัมน์เดียวเริ่มด้วย Key แล้วเรียงคำตอบลงมา จำนวนเฉลยต้องเท่าจำนวนข้อ เลขข้อแบบ Item,Key ต้องเริ่มจาก 1 เรียงครบ ไม่ซ้ำ และไม่มีช่องเฉลยว่าง ไม่ใส่เครื่องหมาย ... เป็นข้อมูลจริง

ระบบตัดช่องว่างหัวท้ายและเปรียบเทียบตัวอักษรแบบไม่แยกตัวพิมพ์ใหญ่เล็ก แต่ไม่ถือว่า A เท่ากับ ก หรือ 1 โดยอัตโนมัติ ต้องใช้รหัสคำตอบชุดเดียวกันทั้งข้อมูลและเฉลย คำตอบว่างหรือไม่ตรงเฉลยถูกให้ 0
ในโหมดคำตอบดิบ จึงควรตรวจว่าความว่างเป็นการไม่ตอบจริงหรือเกิดจากการนำเข้า

ระบุชุดตัวเลือกในช่องตัวเลือกเริ่มต้น เช่น A,B,C,D แล้วใช้กับทุกข้อ หรือแก้เฉพาะข้อในตาราง
เพื่อให้แสดงตัวลวงที่ไม่มีผู้เลือกด้วย

06 ตรวจตัวอย่างข้อมูลและแก้ไขเบื้องต้น

ส่วน “ตรวจสอบข้อมูลก่อนวิเคราะห์” มีตารางตรวจคำตอบ/คะแนน ตารางข้อมูลดิบ และเฉลยกับโครงสร้างข้อสอบ เลือกจำนวนผู้สอบที่แสดงได้ 10 20 50 หรือ 100 คน จำนวนที่แสดงเป็นเพียงตัวอย่าง ไม่ใช่จำนวนผู้สอบที่ใช้วิเคราะห์

เปิด “แก้ไข ID และข้อมูลที่กำลังแสดง” เพื่อแก้ค่าที่เห็น แล้วตรวจความหมายและช่วงคะแนนอีกครั้ง การแก้มีผลต่อรอบวิเคราะห์ปัจจุบันและทำให้ผลเดิมต้องวิเคราะห์ใหม่

การแก้ในหน้าเว็บไม่ได้เขียนกลับไฟล์ต้นฉบับ และ JSON การตั้งค่าไม่เก็บคำตอบทั้งหมด หากต้องการทำซ้ำอย่างตรวจสอบได้ ให้แก้และเก็บไฟล์ต้นฉบับภายนอกด้วย การเปลี่ยนรูปแบบนำเข้าหรือเรียกคืน JSON อาจอ่านตารางต้นทางใหม่ จึงควรหลีกเลี่ยงการใช้ preview เป็นที่เก็บข้อมูลหลัก

ตรวจเฉลยของข้อแรกและข้อท้าย คะแนนเต็มทุกข้อ รหัสที่ไม่ควรเป็นคำว่า ID และหัวข้อที่ไม่ควรถูกนับเป็นผู้สอบ ก่อนเริ่มวิเคราะห์จริง

07 บันทึกและเรียกคืน JSON

ใช้ “บันทึกการตั้งค่าเป็น JSON” หลังตรวจข้อมูลแล้ว ไฟล์เก็บข้อมูลรายวิชา รูปแบบนำเข้า เฉลย คะแนนเต็ม ตัวเลือก มิติย่อย Bloom เป้าหมาย Blueprint เกณฑ์การจัดชุด และวิธีวิเคราะห์รวมถึงที่มาของจุดตัด ไม่ใช่ไฟล์สำรองคำตอบหรือผลวิเคราะห์ทั้งหมด

กรณีเริ่มจากแม่แบบ ให้นำเข้า JSON ก่อน แล้วเลือกไฟล์ข้อมูล ระบบใช้รูปแบบที่ JSON ระบุและตรวจความสอดคล้อง กรณีมีข้อมูลอยู่แล้วสามารถนำเข้า JSON ภายหลังได้ แต่ต้องตรวจจำนวนและ preview หลังรูปแบบถูกนำมาใช้ใหม่

ชื่อไฟล์เหมือนเดิมไม่ได้แปลว่าข้อมูลเหมือนเดิม ระบบตรวจโครงสร้าง ชื่อข้อ และลายนิ้วมือข้อมูล ถ้าข้อมูลตรงจึงเรียกคืนได้โดยตรง หากโครงสร้างตรงแต่คำตอบหรือจำนวนคนเปลี่ยน จะถามยืนยันการใช้การตั้งค่าเดิม

หากจำนวนคนและจำนวนข้อเท่ากันแต่ชื่อข้อหรือรูปแบบบางอย่างต่างกัน อาจใช้เป็นแม่แบบตามลำดับข้อได้เมื่อยืนยัน ผู้ใช้ต้องตรวจเองว่าลำดับและเนื้อหาข้อตรงกันจริง หากจำนวนข้อไม่ตรงหรือจับคู่ไม่ปลอดภัย ระบบจะไม่ใช้การตั้งค่าให้เงียบ ๆ

JSON รุ่นเก่าบางไฟล์อาจบันทึกจากข้อมูลที่เคยถูกตัดผู้สอบหรือหัวข้อผิด อย่ายืนยันใช้เพียงเพื่อให้ผ่านคำเตือน ให้นำเข้าต้นฉบับที่ครบ ตรวจโครงสร้าง แก้ค่าที่จำเป็น แล้วบันทึก JSON ใหม่

แม่แบบ Printout บันทึกแยกในเบราว์เซอร์ ไม่รวมใน JSON การวิเคราะห์ ไม่ควรคาดหวังว่าไฟล์ JSON นี้จะย้ายประวัติการเลือกข้อหรือรูปแบบ Printout ไปอีกเครื่องครบทั้งหมด

08 กำหนดมิติย่อยและระดับ Bloom

กรอกจำนวนตัวชี้วัด/จุดประสงค์ที่แยกวิเคราะห์ได้ตั้งแต่ 0 ถึง 30 แล้วใช้ “ปรับจำนวนมิติ” ตั้งชื่อให้ชัดและไม่ซ้ำ เลือกมิติให้แต่ละข้อจากรายการ ไม่ใช้ชื่อกว้างจนสื่อว่าเป็นมาตรวัดเดียวกันโดยไม่มีหลักฐาน

ค่า “รวมทั้งฉบับ ไม่แยกมิติ” หมายถึงข้อยังรวมคะแนนทั้งฉบับ แต่ไม่สร้างกลุ่มย่อยให้ข้อนั้น มิใช่การตัดข้อออกจากการวิเคราะห์

Bloom มี จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ การกำหนดระดับเป็นการพิจารณาเชิงเนื้อหาของผู้ใช้ ระบบไม่ได้อ่านข้อความข้อสอบแล้วจำแนก Bloom อัตโนมัติ

ปุ่มกำหนดทั้งฉบับเป็นตัวชี้วัดที่เลือก และกำหนด Bloom ให้ทุกข้อ ใช้ลดการกรอกซ้ำ ตรวจตัวเลือกและผลในทุกข้อก่อนยืนยันหรือใช้งานต่อ เพราะเป็นการเปลี่ยนหลายข้อพร้อมกัน

รหัส M1 M2 และสีใช้บอกกลุ่มมิติ ส่วน B1 ถึง B6 ในสัญลักษณ์การจัดกลุ่มหมายถึง Bloom ไม่ใช่ดัชนีจำแนก B อ่านบริบทหัวตารางให้ชัด ชี้เมาส์ โฟกัส หรือเลือกสัญลักษณ์เพื่ออ่านรายละเอียดได้

คะแนนและแอลฟารายมิติขึ้นกับเฉพาะข้อในมิตินั้น หากมีเพียง 1 ข้อหรือคะแนนรวมไม่มีความแปรปรวน แอลฟารายมิติแสดงไม่ได้ การแยกชื่อมิติไม่ได้เป็นหลักฐานว่าโครงสร้างมิติได้รับการยืนยันแล้ว

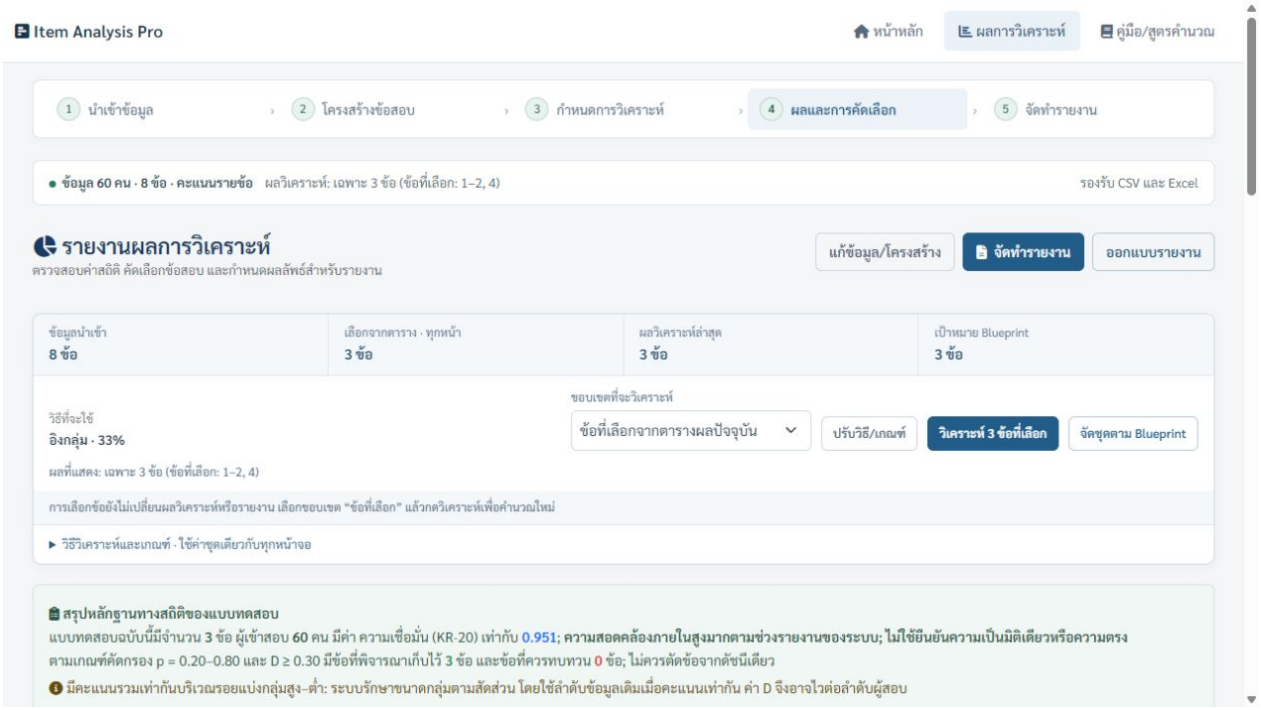
09 แถบวิเคราะห์และความหมายของขอบเขต

ข้อมูลบนแถบ	ความหมาย
ข้อมูลนำเข้า	จำนวนข้อทั้งหมดในไฟล์หลังแยกหัวข้อและรหัส
เลือกจากตาราง ทุกหน้า	จำนวนข้อที่ติ๊กไว้จากผลปัจจุบัน รวมข้อที่อยู่หน้าอื่น
ผลวิเคราะห์ล่าสุด	จำนวนข้อที่ใช้คำนวณผลซึ่งกำลังแสดงและนำไปออกรายงาน
เป้าหมาย Blueprint	จำนวนข้อที่ต้องการจัด ไม่ใช่จำนวนที่เลือกได้จริง

เลือก “ทั้งฉบับจากข้อมูลนำเข้า” เมื่อต้องการวิเคราะห์ทุกข้อ แม้ผลปัจจุบันจะเหลือเพียงชุดย่อย เลือก “ข้อที่เลือกจากตารางผลปัจจุบัน” เมื่อต้องการคำนวณชุดที่ติ๊กไว้ หรือใช้ปุ่มลัด “วิเคราะห์ข้อที่เลือก” ใกล้ตาราง

การติ๊กเลือก การกรองแถว และการซ่อนคอลัมน์ไม่เปลี่ยนผลสถิติเอง ต้องกดวิเคราะห์และตรวจ modal ยืนยันขอบเขต จำนวนคน จำนวนข้อ และจุดตัดก่อน ควรมีผู้สอบและข้อสอบอย่างน้อยอย่างละ 2 แต่การคำนวณได้ไม่แปลว่าขนาดกลุ่มเพียงพอสำหรับข้อสรุปที่มั่นคง

เมื่อเปลี่ยนวิธี เกณฑ์ เฉลย คะแนนเต็ม หรือข้อมูล ผลเดิมจะมีสถานะเตือนและยังไม่ควรนำไปทำรายงานจนวิเคราะห์ใหม่ การวิเคราะห์ใช้ข้อมูลและการตั้งค่าปัจจุบัน ไม่ใช้คะแนนค้างจากรอบก่อน



หน้าจอแสดงขอบเขตผลวิเคราะห์ ข้อมูลสังเคราะห์

10 วิเคราะห์แบบอิงกลุ่ม

เลือก “อิงกลุ่ม” และวิธีแบ่งสูงต่ำ 27% 33% หรือ 50% ระบบเรียงคะแนนรวมของชุดที่จะวิเคราะห์ แล้วใช้จำนวนคนต่อกลุ่มเท่ากับส่วนจำนวนเต็มของ N คูณสัดส่วน โดยไม่น้อยกว่า 1 คน เทคนิค 33% ในโปรแกรมใช้ .33 ไม่ใช่เศษหนึ่งส่วนสามแบบไม่ปัด

ค่า p ในโหมดนี้คำนวณจากค่าเฉลี่ยของกลุ่มสูงและต่ำเทียบคะแนนเต็ม ไม่ใช่สัดส่วนตอบถูกของผู้สอบทุกคนที่นำเข้า ส่วน D เป็นความต่างค่าเฉลี่ยกลุ่มสูงกับต่ำเทียบคะแนนเต็ม ต้องรายงานฐานนี้เมื่อนำไปใช้ในงานวิจัย

ระบบคัดกรองเบื้องต้นด้วย p อยู่ระหว่าง .20 ถึง .80 และ D ไม่น้อยกว่า .30 โดยใช้ค่าก่อนปัดเศษ เกณฑ์นี้เป็นนโยบายคัดกรองของระบบ ไม่ใช่คำตัดสินว่าข้อที่นอกช่วงต้องตัดทิ้งทุกบริบท

เมื่อคะแนนเท่ากันที่รอยแบ่ง ระบบรักษาขนาดกลุ่มโดยใช้ลำดับข้อมูลเดิม จะแจ้งว่าผล D อาจไวต่อลำดับผู้สอบ ควรทบทวนกรณีคะแนนซ้ำมาก กลุ่มเล็ก หรือใช้ผลในการตัดสินใจสำคัญ

หลังลดจำนวนข้อ คะแนนรวมและสมาชิกกลุ่มสูงต่ำอาจเปลี่ยน จึงต้องคำนวณ D และ p ฐานสูงต่ำใหม่ ไม่ควรคัดลอกค่ารายข้อจากฉบับเต็มมาอ้างเป็นผลชุดใหม่โดยไม่วิเคราะห์ซ้ำ

11 วิเคราะห์แบบอิงเกณฑ์

เลือก “อิงเกณฑ์” แล้วกำหนดจุดตัดผ่านของผู้สอบ แยกจากเกณฑ์คัดกรองข้อสอบ เลือกหน่วยร้อยละหรือคะแนนทั้งฉบับ พร้อมบันทึกเหตุผลและวิธีกำหนดจุดตัดตามจุดประสงค์การใช้คะแนน (AERA et al., 2014, Standard 5.21)

ร้อยละ 60 เป็นเพียงค่าตัวอย่าง ไม่ใช่มาตรฐานสากล หากเลือกข้อสอบเพียงบางส่วน
ระบบใช้สัดส่วนเดิมเทียบคะแนนเต็มชุดย่อย เช่น ทั้งฉบับเต็ม 40 ตั้งจุดตัด 24 และชุดย่อยเต็ม 20 จะได้จุดตัด 12 การเทียบนี้ไม่ใช่หลักฐานว่าชุดย่อยมีมาตรฐานผ่านเทียบเท่าฉบับเดิม

ผู้สอบที่ได้คะแนนรวมไม่น้อยกว่าจุดตัดอยู่กลุ่มผ่าน และต่ำกว่าจุดตัดอยู่กลุ่มไม่ผ่าน ค่า p ใช้ผู้สอบทั้งหมด ส่น B สำหรับคะแนน 0/1 เป็นความต่างสัดส่วนตอบถูกระหว่างสองกลุ่ม (Brennan, 1972, 1974)

คะแนนหลายระดับแสดง B-style ซึ่งเป็นความต่างคะแนนเฉลี่ยของสองกลุ่มเทียบคะแนนเต็ม ไม่ควรเรียกเป็น B-index แบบตอบถูกผิดโดยไม่ระบุการขยายวิธีคำนวณ

ค่าเริ่มต้น “บรรยายผลและแจ้งทบทวน” ไม่ใช้ p หรือ B ตัดข้อและไม่เลือกข้อให้อัตโนมัติ
ข้อที่ทุกคนทำได้อาจยังมีประโยชน์ต่อการวัดความรอบรู้ ให้พิจารณาจุดประสงค์และหลักฐานการเรียนรู้ร่วมด้วย

หากเลือก “คัดกรองด้วย B/B-style ขั้นต่ำที่กำหนด” ให้ระบุค่าที่มีเหตุผลรองรับ ค่า .20
เป็นค่าเริ่มต้นของตัวเลือก ไม่ใช่จุดตัดสากล และไม่ใช้ระดับ D เช่น ดีมากหรือชายขอบไปตีความ B โดยอัตโนมัติ

หากไม่มีคนในกลุ่มใดกลุ่มหนึ่ง B คำนวณไม่ได้ แต่คะแนนและ p ยังรายงานได้
ไม่ควรเลื่อนจุดตัดเพียงเพื่อให้เกิดทั้งสองกลุ่มหรือเพื่อเพิ่ม B
การแบ่งกลุ่มในระบบใช้คะแนนรวมที่ยังรวมคะแนนข้อนั้น
ไม่ใช่เกณฑ์ภายนอกหรือหลักฐานความไวต่อการสอนโดยลำพัง

12 อ่านผลสรุป ตาราง และกราฟ

หน้าผลแสดงจำนวนข้อ ค่าเฉลี่ย ความเชื่อมั่น SEM จำนวนผู้สอบ SD KR-21 และวิธีวิเคราะห์ ตรวจชื่อสัมประสิทธิ์ก่อนอ่านค่า อิงกลุ่มใช้ KR-20 เมื่อทุกข้อเป็น 0/1 หรือ Alpha เมื่อเป็นคะแนนหลายระดับ อิงเกณฑ์แสดง Livingston k² และเก็บค่าความสอดคล้องภายในไว้ในรายละเอียดรายงาน

แท็บ “วิเคราะห์รายข้อ” แสดง p การแปลผล D/B สหสัมพันธ์ item-rest Alpha เมื่อตัดข้อ Modified Cj* และข้อเสนอแนะ คลิกปุ่มรายละเอียดเพื่ออ่านชื่อมิติ Bloom ค่าดัชนี และบริบทครบถ้วน

แท็บ “วิเคราะห์มิติย่อย” แสดงจำนวนข้อ คะแนนเต็ม ค่าเฉลี่ย แอลฟา และรายการข้อในแต่ละมิติ ส่วน “ผลรายผู้เข้าสอบ” แสดงคะแนนรวม สถานะผ่านเมื่อใช้อิงเกณฑ์ ดัชนีรูปแบบตอบเมื่อคำนวณได้ และคะแนนรายมิติ

เลือก “แบบกระชับ” หรือ “ข้อความเต็ม” เพื่อจัดความหนาแน่นของตาราง ใช้คำอธิบายรหัสเพื่อแยก M ของมิติ B ของ Bloom และ B-index ในผลจำแนก สีไม่ได้เป็นระดับคุณภาพ

ตารางรองรับเรียงลำดับ ค้นหา เปลี่ยนจำนวนข้อต่อหน้า กรองมิติและ Bloom และเลือกคอลัมน์ที่แสดง การซ่อนคอลัมน์บนหน้าจอไม่ใช่คำสั่งลบคอลัมน์จากรายงาน ซึ่งต้องกำหนดในตัวออกแบบรายงานแยกต่างหาก

กราฟการแจกแจงคะแนนและแผนภาพ p กับอำนาจจำแนกอยู่ส่วนพับหลังตาราง ไม่ใช่การทดสอบนัยสำคัญหรือยืนยันสมมติฐานการแจกแจง เครื่องหมายขีดหมายถึงคำนวณไม่ได้หรือไม่เข้าเงื่อนไข ไม่ใช่ค่าศูนย์

13 เลือกอัตโนมัติ เลือกเอง และคำนวณใหม่

“เลือกตามเกณฑ์คัดกรอง” แทนรายการเลือกปัจจุบันด้วยข้อที่เข้าเงื่อนไขจากผลล่าสุด
หากมีการเลือกเองมาก่อนควรบันทึกรายการไว้ก่อนกดซ้ำ ในโหมดอิงเกณฑ์แบบบรรยาย ปุ่มเปลี่ยนเป็น “เลือกข้อทั้งหมด ไม่คัดกรอง” เพื่อไม่อ้างว่าผ่านเกณฑ์ที่ไม่ได้ใช้

“เพิ่มข้อจากผลกรอง” เพิ่มข้อที่ตรงตัวกรองเข้ารายการเดิม ไม่ล้างข้อที่เลือกหน้าอื่น “ยกเลิกการเลือกทั้งหมด” ล้างรายการ ส่วนช่องเลือกทั้งหมดในหัวตารางหมายถึงทุกข้อของผลปัจจุบัน ไม่ใช่เพียงหน้าที่มองเห็น

เลือกเองได้โดยติ๊กหรือยกเลิกแต่ละข้อ ระบบคงการเลือกข้ามการเรียงลำดับ การกรอง และการเปลี่ยนหน้า จากนั้นกด “วิเคราะห์ข้อที่เลือก” ตรวจเลขข้อเดิมใน modal แล้วยืนยัน

หลังวิเคราะห์ใหม่ ระบบคงรายการที่ยังอยู่ในขอบเขตใหม่ไว้ เพื่อปรับเกณฑ์และวิเคราะห์ซ้ำได้
หากต้องเพิ่มข้อที่ถูกตัดออกจากผลชุดย่อย ให้กลับมาวิเคราะห์ “ทั้งฉบับจากข้อมูลนำเข้า” ก่อน แล้วจึงเลือกใหม่

ค่าสถิติ	หลังเปลี่ยนชุดข้อสอบ
จำนวนคน N	คงเดิมหากใช้ผู้สอบกลุ่มเดิม
คะแนนรวม ค่าเฉลี่ย SD	คำนวณจากคะแนนชุดใหม่
Alpha KR-20 KR-21 SEM	คำนวณใหม่เมื่อมีเงื่อนไขให้คำนวณ
D และ p แบบสูงต่ำ	อาจเปลี่ยนเพราะอันดับและกลุ่มเปลี่ยน
B และสถานะผ่าน	อาจเปลี่ยนเพราะคะแนนรวมและจุดตัดชุดใหม่
p อิงเกณฑ์รายข้อ	อาจคงเดิมเมื่อผู้สอบและคะแนนข้อนั้นเหมือนเดิม
item-rest และ Alpha เมื่อตัดข้อ	คำนวณเทียบกับข้อที่เหลือในชุดใหม่

ตัวอย่างตรวจระบบด้วยข้อมูลสังเคราะห์ 60 คน: ทั้งฉบับ 8 ข้อได้ค่าเฉลี่ย 4.72 และ KR-20 .594
เมื่อเลือกตามเกณฑ์แล้ววิเคราะห์ 6 ข้อ ได้ 3.22 และ .884 เมื่อเลือกเองข้อ 1 2 และ 4 แล้ววิเคราะห์ 3 ข้อ ได้ 1.50 และ .951 ตัวเลขนี้ใช้สาธิตการคำนวณใหม่ ไม่ใช่เกณฑ์รับรองคุณภาพ และค่าสูงอาจเกิดจากข้อซ้ำซ้อน

14 จัดชุดตาม Blueprint

เปิด “จัดชุดตาม Blueprint” ระบุจำนวนข้อในแต่ละจุดประสงค์คูณระดับ Bloom เป้าหมายต้องเป็นจำนวนข้อไม่ติดลบ และมีอย่างน้อยหนึ่งกลุ่มที่ต้องการข้อ กำหนดมิติและ Bloom ของข้อในหน้าโครงสร้างให้ครบก่อน

เลือก “ใช้ตามนโยบายผลวิเคราะห์” เพื่อสืบทอดแนวพิจารณาปัจจุบัน อิงกลุ่มใช้ p .20-.80 และ D อย่างน้อย .30 อิงเกณฑ์แบบบรรยายไม่ตัดจาก p/B ส่วนโหมดคัดกรองใช้ B ขั้นต่ำที่วิเคราะห์ไว้ หากเลือก “กำหนดเกณฑ์สถิติเฉพาะการจัดชุด” ให้ตรวจ D/B และช่วง p ที่กำหนดเอง

กด “คัดเลือกชุดข้อสอบ” ระบบเลือกข้อไม่ซ้ำภายในแต่ละกลุ่มจากขอบเขตผลปัจจุบัน โหมดไม่คัดกรองเสนอเรียงเลขข้อเดิม โหมดคัดกรองอิงเกณฑ์เรียง B มากไปน้อย ส่วนอิงกลุ่มใช้คะแนนจัดอันดับเชิงปฏิบัติ .60D + .20r item-rest + .20[1 - 2|p - .50|]

สูตรจัดอันดับเป็นวิธีเสนอรายการของระบบ ไม่ใช่การหาชุดที่เหมาะสมที่สุดโดยมีหลักประกัน ไม่ได้เพิ่มความเชื่อมั่นหรือยืนยันความตรงโดยอัตโนมัติ ข้อที่ไม่ระบุ Bloom ไม่ถูกนำเข้ากลุ่มเป้าหมาย

ตรวจคอลัมน์ต้องการ จำนวนที่มีหรือผ่านคัดกรอง รายการข้อ และสถานะ หากกลุ่มใดไม่ครบ ระบบแจ้งจำนวนที่ขาด ควรเพิ่มคลัง ตรวจการจัดกลุ่ม หรือทบทวนเงื่อนไขอย่างมีเหตุผล ไม่ลดเกณฑ์เพียงเพื่อให้ครบโดยไม่บันทึกเหตุผล

“นำรายการไปปรับเลือกเอง” ส่งรายการข้อที่เสนอไปยังช่องเลือกในตาราง เพื่อเพิ่มหรือลดเองก่อนวิเคราะห์ “วิเคราะห์ชุดที่จัดใหม่” ส่งรายการนั้นเข้าสู่การยืนยันและคำนวณคะแนนกับดัชนีของชุดใหม่โดยตรง หากมีไม่ถึง 2 ข้อ ระบบยังไม่คำนวณความเชื่อมั่นให้

รายงาน Blueprint เป็นหลักฐานรายการคัดเลือก โดยค่าสถิติรายข้ออ้างอิงขอบเขตต้นทางที่ระบุ ไม่ใช่ผลวิเคราะห์ใหม่ของชุดที่จัด หากต้องการผลหลังคัดเลือก ให้สั่งวิเคราะห์ชุดใหม่ แล้วจัดทำรายงานผลวิเคราะห์แยก เมื่อวิเคราะห์ใหม่ผลการจัดชุดเดิมจะถูกยกเลิกเพื่อไม่ให้สับสนกับข้อมูลคนละรอบ

15 วิเคราะห์ตัวเลือก ตัวลวง และ Trace Line

ใช้ได้เมื่อมีคำตอบรายตัวเลือกพร้อมเฉลย ไม่สามารถสร้างการแจกแจงตัวลวงกลับจากคะแนน 0/1 ที่ไม่มีคำตอบดิบได้ เปิด “วิเคราะห์ตัวลวงเชิงลึก” ในแท็บรายข้อ แล้วตรวจทุกตัวเลือก รวมตัวเลือกที่ไม่มีผู้เลือกตามชุดตัวเลือกที่กำหนด

แสดงจำนวนและร้อยละผู้เลือกทั้งกลุ่ม พร้อมร้อยละในกลุ่มสูง H และต่ำ L การจัดกลุ่มส่วนนี้ใช้คะแนนรวมที่ตัดคะแนนข้อนั้นออก และสัดส่วน 27/33/50 ที่ตั้งไว้ จึงอาจไม่ตรงกับ D รายข้อในตารางหลักที่ใช้คะแนนรวมรวมข้อนั้น

สำหรับตัวถูก พิจารณาสัดส่วน H ลบ L ส่วนตัวลวงพิจารณา L ลบ H ตัวลวงที่มีผู้เลือกน้อยกว่า 5% หรือมีทิศทางกลับควรตรวจเนื้อหา แต่เกณฑ์ความถี่ไม่สามารถแทนการพิจารณาความสมเหตุสมผลของตัวลวงได้ (Haladyna & Downing, 1993; Tarrant et al., 2009)

แม้วิเคราะห์แบบอิงเกณฑ์ การแจกแจงตัวลวงในเครื่องมือนี้ยังเป็นการสำรวจตามกลุ่มคะแนนสูงต่ำ ไม่ใช่การเปรียบเทียบผ่านกับไม่ผ่านโดยใช้จุดตัด ต้องระบุฐานวิเคราะห์เมื่อเขียนรายงาน

เลือกข้อในช่อง Trace Line เพื่อดูร้อยละเลือกแต่ละตัวเลือกใน 4 กลุ่มคะแนนจากต่ำไปสูง โดยใช้คะแนนที่ตัดข้อนั้นออก กลุ่มสุดท้ายรับจำนวนที่เหลือ การแบ่งกลุ่มมีผลเมื่อ N น้อยหรือคะแนนเท่ากัน จึงควรใช้ประกอบการตรวจทานมากกว่าตัดสินจากรูปร่างกราฟอย่างเดียว

16 จัดทำรายงานมาตรฐานและส่งออก

กด “จัดทำรายงาน” แล้วเลือกชุดรายงานทางวิชาการ ทบทวนคุณภาพข้อสอบ หรือรายงานพร้อมภาคผนวก สามารถปรับเลือกแต่ละองค์ประกอบเพิ่มหรือลดได้อีก ไม่จำเป็นต้องใช้ตาม preset ทุกส่วน

มีองค์ประกอบ 9 ส่วน คือสรุปผล Blueprint และ Bloom ผลรายข้อ ผลแยกมิติ ข้อเสนอแนะ เกณฑ์กับเอกสารอ้างอิง ผลรายผู้สอบ ภาคผนวกตัวลวง และภาคผนวกคำตอบดิบ สองส่วนท้ายต้องมีข้อมูลคำตอบดิบ

ตรวจข้อความขอบเขตผลในหน้าต่างรายงาน หากรายการที่เลือกไว้มีจำนวนต่างจากผลล่าสุด ระบบเตือนว่ารายงานยังใช้ผลล่าสุด การเลือกรายการข้อไว้เฉย ๆ ไม่เปลี่ยนค่าสถิติหรือขอบเขตรายงาน

กำหนดชื่อรายงาน A4 แนวตั้งหรือแนวนอน และระยะตาราง แล้วเปิดหน้าตรวจสอบ HTML ในแท็บใหม่ หากเบราว์เซอร์ระงับหน้าต่างใหม่ ให้ใช้ลิงก์หน้ารายงานที่ระบบเตรียมไว้หรืออนุญาตหน้าต่างสำหรับเว็บไซต์นี้ตามนโยบายของหน่วยงาน

ในหน้ารายงาน เลือกส่วนที่จะใช้ เลือกทั้งหมดหรือยกเลิกทั้งหมด ปรับกระดาษ แนววาง ขนาดอักษร ระยะตาราง หมายเหตุ การแยกส่วนขึ้นหน้าใหม่ และขาวดำได้ ส่วนที่ไม่ได้เลือกไม่ถูกนำไปส่งออก ถ้าไม่เลือกเลยระบบปิดการส่งออก

“PDF / พิมพ์” เปิดหน้าต่างพิมพ์ของเบราว์เซอร์ ให้เลือกบันทึกเป็น PDF ตรวจขนาดกระดาษ สเกล และหัวท้ายกระดาษของเบราว์เซอร์ก่อนบันทึก ไม่ใช่การดาวน์โหลด PDF โดยไม่ผ่านหน้าต่างพิมพ์

“Word” ส่งออก .doc ที่ใช้โครงสร้าง HTML ไม่ใช่ .docx ดั้งเดิม Word อาจแจ้งว่ารูปแบบไม่ตรงนามสกุล ให้เปิดเฉพาะไฟล์ที่สร้างจากระบบของตน แล้วตรวจหน้าและบันทึกเป็น .docx หากต้องแก้ไขต่อ การแบ่งหน้าอาจต่างจาก HTML

“HTML” บันทึกหน้ารายงานพร้อมส่วนควบคุมที่คงตัวเลือกไว้ ใช้ TH Sarabun New เมื่อมีในเครื่องและฟอนต์สำรองเมื่อไม่มี ตรวจภาษาไทยและหน้าสุดท้ายทุกครั้ง คำสั่งคัดลอกทั้งส่วนที่เลือกหรือคัดลอกเฉพาะส่วนใช้เพื่อนำตารางไปวางในรูปเล่มได้

17 ออกแบบ Printout ให้เหมาะกับงาน

เข้า “ออกแบบรายงาน” จากหน้าผล หรือ “กำหนดหัวตารางและลำดับส่วนรายงาน” ในศูนย์รายงาน
เครื่องมือนี้ใช้ขอบเขตผลวิเคราะห์ล่าสุด ไม่ได้วิเคราะห์ข้อใหม่ให้เอง

กำหนดชื่อไทย ชื่อรอง ข้อความนำ ท้ายรายงาน แนวกระดาษ ขนาดกระดาษ ระยะตาราง รูปแบบสี
และข้อมูลหัวรายงานที่ต้องการ เช่น รายวิชา ผู้รับผิดชอบ วันที่ และขอบเขต

ลากส่วนรายงานเพื่อเปลี่ยนลำดับ หรือใช้ปุ่มขึ้นลงสำหรับการใช้งานที่ไม่สะดวกลาก
เลือกเปิดปิดส่วนตามต้องการ การลากเป็นการจัดลำดับส่วนรายงาน
ไม่ใช่การลากเซลล์อิสระเหมือนโปรแกรมจัดหน้า

เลือกคอลัมน์ตารางรายข้อและแก้ชื่อหัวคอลัมน์ได้ รวม 16 ตัวเลือก เช่น ข้อ มิติ Bloom คะแนนเต็ม p
เปอร์เซ็นต์กลุ่มผ่าน/ไม่ผ่าน D/B item-rest Alpha เมื่อตัดข้อ Cj* คำแปลผล และข้อเสนอแนะ
คอลัมน์ที่เปิดมากอาจต้องใช้แนวนอนหรือแยกเนื้อหา ไม่ควรย่อตัวอักษรจนอ่านไม่ได้

ตัวเลือกหัวตารางซ้ำ เลขหน้า และข้อความความเป็นส่วนตัวขึ้นกับความสามารถของตัวส่งออกและเบราว์เซอร์
ต้องตรวจผลพิมพ์จริง ไม่ถือว่าการเปิดตัวเลือกยืนยันว่าทุกโปรแกรมจะแบ่งหน้าเหมือนกัน

บันทึกแม่แบบเพื่อเรียกใช้ในเบราว์เซอร์และโดเมนเดิมได้ การล้างข้อมูลเว็บไซต์หรือย้ายเครื่องอาจทำให้แม่แบบหาย ปุ่มคืนค่าเริ่มต้นคืนเฉพาะรูปแบบ ไม่ใช่การกู้คืนคำตอบหรือผลวิเคราะห์

เปิด HTML preview ตรวจหัวตาราง รหัสย่อกับคำอธิบาย และเลขข้อเดิมก่อนส่งออก
การเปลี่ยนชื่อคอลัมน์เพื่อความกระชับไม่ควรเปลี่ยนความหมายทางวิชาการของดัชนี

18 สูตรและฐานคำนวณที่ควรทราบ

กำหนด N เป็นจำนวนผู้สอบ k เป็นจำนวนข้อ X เป็นคะแนนรวม x เป็นคะแนนรายข้อ m เป็นคะแนนเต็มรายข้อ และ c เป็นจุดตัด สูตรในบทนี้สรุปการคำนวณของรุ่นนี้ ไม่ได้หมายความว่าทุกทางเลือกเป็นมาตรฐานที่ต้องใช้ในทุกงาน

ค่าเฉลี่ยเท่ากับผลรวม X หาร N ความแปรปรวนของคะแนนที่ระบบใช้หารด้วย N และ SD
เป็นรากที่สองของค่านั้น จึงอาจต่างจาก sample SD ที่ใช้ N - 1 ในโปรแกรมอื่น
ต้องใช้ฐานเดียวกันเมื่อตรวจเทียบตัวเลข

p อิงกลุ่มเท่ากับผลบวกค่าเฉลี่ยข้อในกลุ่มสูงกับต่ำ หารด้วย 2m ส่วน D เท่ากับค่าเฉลี่ยข้อกลุ่มสูงลบต่ำ หารด้วย m สำหรับคะแนน 0/1 เทียบได้กับการใช้สัดส่วนตอบถูกของแต่ละกลุ่ม

p อิงเกณฑ์เท่ากับคะแนนเฉลี่ยข้อนั้นของผู้สอบทั้งหมดหาร m ส่วน B/B-style
เท่ากับค่าเฉลี่ยข้อของกลุ่มผ่านลบกลุ่มไม่ผ่าน หาร m หากกลุ่มใดว่างจะไม่นิยามดัชนีความต่างระหว่างกลุ่ม

item-rest เป็นสหสัมพันธ์ Pearson ระหว่างคะแนนข้อกับคะแนนรวมที่ตัดข้อนั้นออก
คำนวณไม่ได้หากคะแนนตัวแปรใดไม่มีความแปรปรวน สำหรับคะแนนสองค่าเกี่ยวข้องกับ point-biserial
แต่สำหรับคะแนนหลายระดับควรรายงานเป็นสหสัมพันธ์ item-rest

Alpha = [k / (k - 1)] × [1 - ผลรวมความแปรปรวนรายข้อ / ความแปรปรวนคะแนนรวม] สำหรับข้อมูล 0/1 ทุกข้อสอดคล้องกับ KR-20 ส่วน KR-21 ใช้สูตรย่อจาก k ค่าเฉลี่ยและความแปรปรวนคะแนนรวม
จึงมีสมมติฐานเข้มกว่าและไม่ควรใช้แทน KR-20 โดยไม่พิจารณา

Alpha เมื่อตัดข้อคำนวณจากข้อที่เหลือจริง ระบบไม่แสดงเมื่อเหลือข้อไม่เพียงพอหรือคะแนนไม่มีความแปรปรวน ค่าติดลบเป็นสัญญาณต้องทบทวน ไม่ควรแก้เป็นศูนย์เพื่อให้รายงานดูดี

SEM = SD × รากที่สองของ (1 - r) โดยใช้ความสอดคล้องภายใน KR-20/Alpha ไม่ใช้ Livingston k² แทน หาก r อยู่นอกช่วง 0 ถึง 1 หรือคำนวณไม่ได้ ระบบไม่แสดง SEM เป็นค่าปกติ

Livingston k² = [r × ความแปรปรวน X + (ค่าเฉลี่ย X - c)²] / [ความแปรปรวน X + (ค่าเฉลี่ย X - c)²]
ต้องอ่านร่วมกับจุดตัดและข้อจำกัด ไม่ใช่อัตราจำแนกถูกต้อง (Livingston, 1972)

Modified Cj* เปรียบเทียบผลรวมคะแนนรวมของผู้ที่ทำข้อนั้นได้จริงกับรูปแบบสูงสุดและต่ำสุดที่เป็นไปได้
ดัชนีรายผู้สอบ Ci* เปรียบเทียบผลรวมความยากตามข้อที่ตอบถูกกับรูปแบบที่สอดคล้องที่สุดและน้อยที่สุด
คำนวณเฉพาะคะแนน 0/1 ทุกข้อ และไม่แสดงกรณีตัวหารเป็นศูนย์หรือได้สุดขอบทั้งหมด
ดัชนีเหล่านี้ใช้ชี้ประเด็นตรวจทาน ไม่ใช้วินิจฉัยการทุจริต (Harnisch & Linn, 1981)

19 เกณฑ์แปลผลและการใช้ในงานวิจัย

ค่า p	ข้อความบรรยายของระบบ
0 ถึง .20	ยากมาก
มากกว่า .20 ถึง .40	ยาก
มากกว่า .40 ถึง .60	ปานกลาง
มากกว่า .60 ถึง .80	ง่าย
มากกว่า .80 ถึง 1	ง่ายมาก

ค่า p สูงแปลว่าผู้สอบทำได้มาก แม้เรียกดัชนีนี้ว่า “ค่าความยาก” ช่วงคำบรรยายเป็นเกณฑ์ปฏิบัติของระบบ ไม่ใช่กฎเก็บหรือตัดข้อ ช่วงคัดกรอง p .20-.80 รวมค่าขอบทั้งสอง จึงอาจมี p = .20 ซึ่งบรรยายยากมากแต่ยังเข้าเงื่อนไขคัดกรองเมื่อ D ถึงเกณฑ์

D ตั้งแต่ .40 บรรยายดีมาก .30 ถึงต่ำกว่า .40 ดี .20 ถึงต่ำกว่า .30 ชายขอบ มากกว่า 0 แต่ต่ำกว่า .20 ต่ำ ศูนย์ไม่พบความต่าง และติดลบทิศทางตรงข้าม ใช้ร่วมกับ p เนื้อหา และวัตถุประสงค์ ไม่โอนช่วงนี้ไปตัดสิน B โดยตรง (Ebel & Frisbie, 1991)

item-rest ที่ .30 เป็นแนวพิจารณาที่ใช้กันในการพัฒนามาตรวัด แต่ไม่ยืนยันความตรง ช่วง .20 และข้อความย่อยในระบบเป็นแนวตรวจทานของโปรแกรม แอลฟาระดับ .70 .80 .90 ต้องใช้ตามบริบทและผลกระทบของการใช้คะแนน ไม่ใช่ข้อกำหนดสากลสำหรับทุกกลุ่มหรือทุกการตัดสินใจ

จุดแจ้ง C* ตั้งแต่ .30 ความถี่ตัวลวง 5% และคำเตือนกลุ่มน้อยกว่า 30 คนหรือกลุ่มย่อยน้อยกว่า 5 คน เป็นเงื่อนไขตรวจทานที่ระบบใช้ ต้องแยกจากแหล่งสูตร ไม่ควรเขียนว่าผู้เสนอสูตรรับรองจุดเหล่านี้ทุกกรณี

รายงานงานวิจัยควรระบุประชากรหรือกลุ่มตัวอย่าง จำนวนคนและข้อ รูปแบบคะแนน วิธีแบ่งกลุ่ม ฐาน p จุดตัดและที่มา นโยบายคัดกรอง สูตรที่ใช้ ชื่อสัมประสิทธิ์ และข้อจำกัด หลีกเลี่ยงคำว่า “ข้อสอบมีคุณภาพแล้ว” จากค่าดัชนีเดียว

ตัวอย่างถ้อยคำอิงเกณฑ์คือ “วิเคราะห์คะแนนจากผู้สอบ ... คน โดยใช้จุดตัด ... จากคะแนนเต็ม ... ตาม ... รายงาน p จากผู้สอบทั้งหมดและ B/B-style ระหว่างกลุ่มผ่านกับไม่ผ่าน ใช้ผลร่วมกับความครอบคลุมตาม Blueprint และการตรวจเนื้อหา” เติมข้อมูลจริงและหลักฐานที่มี ไม่ใช้ข้อความนี้สร้างวิธีกำหนดมาตรฐานที่ไม่ได้ทำ

20 ชุดตัวอย่างสำหรับการสอน

เมนูดาวน์โหลดตัวอย่างมีคะแนน 0/1 แบบมีและไม่มีรหัสอย่างละ 48 คน × 20 ข้อ คะแนนอัตนัย 48 คน × 6 ข้อ คำตอบ A-D 48 คน × 20 ข้อพร้อมไฟล์เฉลย และคะแนนผสม 48 คน × 10 ข้อ

ชุดฝึก Sato มี 36 คน × 20 ข้อเพื่อสังเกตรูปแบบที่ควรตรวจทาน ชุดฝึกมิติย่อยมี 48 คน × 24 ข้อ ให้สร้าง 3 มิติแล้วจัดข้อ 1-8 9-16 และ 17-24 ตามลำดับ ส่วนไฟล์ Error ใช้ฝึกอ่านคำเตือน ไม่ใช่ข้อมูลที่ต้องทำให้วิเคราะห์ผ่าน

ตัวอย่างเป็นข้อมูลสังเคราะห์กำหนดผลซ้ำได้ ออกแบบให้มีทั้งข้อทำงานดีและข้อควรทบทวน ไม่ใช่ผลทดลองกับผู้เรียนจริง การดาวน์โหลดชุดมิติไม่ได้หมายความว่าระบบกำหนดแท็กมิติและ Bloom ให้ทุกข้อแล้ว ต้องตั้งค่าหรือเรียกคืน JSON ให้ตรง

สำหรับชุดคะแนนอัตนัย ตรวจคะแนนเต็มตามการออกแบบ 5 10 5 15 10 และ 20 ตามลำดับ ไม่เปลี่ยนตามคะแนนสูงสุดที่พบโดยไม่มีเหตุผล สำหรับชุดผสมให้ตรวจทั้งส่วน 0/1 และส่วนหลายคะแนนก่อนเปรียบเทียบดัชนี

21 การแจ้งเตือนและการแก้ปัญหา

นำเข้าแล้วจำนวนคนหรือข้อไม่ตรง ให้ตรวจแนวตาราง ตัวเลือกรหัส และตัวเลือกหัวข้อก่อนแก้คะแนน
ระบบไม่ลบหัวข้อเอง หากเผลอนับหัวเป็นผู้สอบ ให้เลือกหัวข้อให้ถูก ไม่ลบผู้สอบคนแรกจากต้นฉบับเพื่อชดเชย

จำนวนคอลัมน์ไม่เท่ากัน ให้ตรวจแถวที่ระบบระบุ ตัวคั่น และเครื่องหมายคำพูด CSV ต้องปิดคำพูดให้ครบ
ถ้าชื่อมี comma ให้ใส่ในเครื่องหมายคำพูด หรือจัดไฟล์ Excel ที่ตารางชัดเจน

คะแนนเป็นข้อความหรือเกินคะแนนเต็ม ให้ตรวจโหมด ตัวเลขทศนิยม และคะแนนเต็มจริง
ถ้าเป็นรหัสตัวเลือกเลขให้เปิดคำตอบดิบ ไม่แก้คะแนนผิดเป็น 0 โดยไม่มีหลักฐาน

เฉลยไม่ครบหรือมีเลขข้อซ้ำ ให้ตรวจว่ามีคำตอบหนึ่งตำแหน่งต่อข้อ ไม่มีช่องว่างหรือหัวข้อเกิน
และเลือกไฟล์เฉลยของแบบทดสอบฉบับเดียวกับคำตอบ

JSON ไม่ตรงข้อมูล ให้ตรวจจำนวนข้อ ลำดับ ชื่อข้อ รูปแบบนำเข้า และโหมดคำตอบ ชื่อไฟล์ตรงกันไม่เพียงพอ
การใช้แม่แบบกับข้อมูลใหม่ต้องยืนยันว่าจับคู่ตามลำดับข้อได้จริง

B คำนวณไม่ได้ ให้ตรวจจำนวนคนผ่านและไม่ผ่าน ไม่เปลี่ยนจุดตัดเพียงเพื่อให้คำนวณได้ Alpha
หรือสหสัมพันธ์เป็นขีด ให้ตรวจจำนวนข้อและความแปรปรวน ไม่แทนขีดด้วยศูนย์

เลือกข้อแล้วค่ายังเท่าเดิม ให้ตรวจว่าได้กดวิเคราะห์ข้อที่เลือกหรือยัง หากกดแล้ว N และ p
อิงเกณฑ์ยังเท่าเดิมอาจถูกต้องตามฐานคำนวณ ให้ดูขอบเขต จำนวนข้อ และค่าสถิติที่ขึ้นกับชุดข้อร่วมกัน

Blueprint ไม่ครบ ให้ตรวจมิติและ Bloom ของข้อ เป้าหมาย และเกณฑ์ที่ใช้
รายงานต้นทางกับผลชุดใหม่เป็นคนละขั้น ไม่ใช้การครบเป้าหมายรับรองคุณภาพ

รายงานไม่เปิด ให้ตรวจการระงับหน้าต่างใหม่ หากปิดตัวเลือกทุกส่วนจะส่งออกไม่ได้ PDF ให้ตรวจหน้าต่างพิมพ์
ส่วน Word ให้ตรวจว่าเปิดไฟล์ .doc แบบ HTML ได้และมีฟอนต์ไทย หลีกเลี่ยงการกดปุ่มซ้ำระหว่างเปิดหน้าต่าง

ภาษาไทยหรือหัวตารางล้น ให้ตรวจฟอนต์ ลดจำนวนคอลัมน์หรือเลือกแนวนอน
ใช้ชื่อย่อพร้อมหมายเหตุแทนข้อความยาวในหัวตาราง ตรวจหน้าสุดท้ายและตารางหลายหน้าก่อนนำไปใช้จริง

22 ติดตั้ง สำรองงาน และตรวจรับ

สำหรับ Hostatom ให้สำรองเว็บไซต์เดิมก่อน แล้วอัปโหลด index.html app.js academic.js ui.css และ ui.js พร้อมโฟลเดอร์ docs สำหรับคู่มือ ไปยังโฟลเดอร์เว็บไซต์ที่โดเมนใช้งานจริง ไม่ต้องอัปโหลด code.gs หรือไฟล์ทดสอบใน tmp และไม่ต้องสร้าง Google Sheet หรือฐานข้อมูลสำหรับรุ่นนี้

เปิดผ่าน HTTPS ตรวจว่าไฟล์ JavaScript CSS และ CDN โหลดได้ ใช้ตัวอย่างสังเคราะห์ทดสอบนำเข้า วิเคราะห์ทั้งฉบับ คัดเลือกชุดย่อย และรายงานบนโดเมนจริง บันทึกงานก่อนรีเฟรชเพื่อรับไฟล์รุ่นใหม่

เก็บไฟล์ข้อมูลต้นฉบับที่แก้ครบ ไฟล์เฉลย JSON การตั้งค่า รายการข้อที่คัดเลือก และรายงานผลของแต่ละรอบแยกกัน ใช้ชื่อที่บอกวิชา รุ่นแบบทดสอบ วันที่ และขอบเขต ไม่ใช้ JSON แทนข้อมูลต้นฉบับ

รายการตรวจรับก่อนใช้ผลจริงคือจำนวนคนและข้อครบ เฉลยและคะแนนเต็มถูก มิติและ Bloom ตรงเนื้อหา วิธีและจุดตัดมีเหตุผล รายงานใช้ขอบเขตล่าสุด ตรวจตารางและภาษาไทยครบทุกหน้า และไม่เปิดเผยข้อมูลผู้สอบเกินจำเป็น

ชุดทดสอบซอฟต์แวร์และการทดสอบข้อมูลสังเคราะห์ช่วยตรวจพฤติกรรมที่ระบุ ไม่ใช่การรับรองทุกไฟล์ ทุกอุปกรณ์ หรือความเหมาะสมทางวิชาการของแบบทดสอบจริง ผู้รับผิดชอบควรตรวจรับบน Hostatom และในโปรแกรมเอกสารที่หน่วยงานใช้ก่อนเผยแพร่ผลสำคัญ

23 เอกสารอ้างอิง

รายการต่อไปนี้ใช้ร่วมกับคู่มือเกณฑ์ในเว็บไซต์ แยกแหล่งทฤษฎีออกจากนโยบายคัดกรองของโปรแกรม การอ้างอิงงานต้นฉบับไม่ได้รับรองว่าค่าตัดของระบบเหมาะสมกับทุกบริบท

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Brennan, R. L. (1972). A generalized upper-lower item discrimination index. *Educational and Psychological Measurement, 32*(2), 289–303. https://doi.org/10.1177/001316447203200206

Brennan, R. L. (1974). *The evaluation of mastery test items: Final report*. State University of New York at Stony Brook. https://files.eric.ed.gov/fulltext/ED092593.pdf

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement, 53*(4), 999–1010. https://doi.org/10.1177/0013164493053004013

Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement, 18*(3), 133–146. https://doi.org/10.1111/j.1745-3984.1981.tb00848.x

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160. https://doi.org/10.1007/BF02288391

Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. *Journal of Educational Measurement, 9*(1), 13–26. https://doi.org/10.1111/j.1745-3984.1972.tb00756.x

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education, 9*, Article 40. https://doi.org/10.1186/1472-6920-9-40