

บทที่ 3

วิธีการวิจัย

การวิจัยเรื่อง “การสร้างและพัฒนาแบบทดสอบความสามารถทั่วไปทางสมองที่วัดทั้งความสามารถและความรวดเร็วในการตอบ สำหรับข้าราชการพลเรือนไทยระดับแรกบรรจุ” ประกอบด้วยหัวข้อในการดำเนินการ ดังนี้

1. ประชากร
2. กลุ่มตัวอย่าง
3. เครื่องมือที่ใช้ในการวิจัย
4. วิธีการวิจัย
5. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ประชากร

ประชากร คือ กลุ่มข้าราชการพลเรือนไทยในสายงานระดับแรกบรรจุ ระดับ 3 หรือ 4 ซึ่งยังดำรงตำแหน่งอยู่ในระดับ 3-5 จากส่วนราชการระดับกรม ในเขตกรุงเทพมหานครและปริมณฑล จำนวน 1,300 คน รวม 143 สายงาน จากกลุ่มสายงาน 8 กลุ่ม ตามมาตรฐานกำหนดตำแหน่ง (สำนักงาน ก.พ., 2548, น. 86-96) (ดังภาคผนวก ก) ทั้งนี้ ไม่รวมสายงานที่ปัจจุบันไม่มีข้าราชการในระดับ 3-5 (สำนักงาน ก.พ., 2548, น. 86-96) และไม่รวมสายงานที่สามารถสมัครได้เฉพาะผู้สำเร็จการศึกษาในวุฒิปริญญาตรี ก.พ. กำหนดให้คัดเลือกเพื่อบรรจุบุคคลเข้ารับราชการ (ตามหนังสือเวียนสำนักงาน ก.พ. ที่ นร. 1004.1/ว. 16) เนื่องจากเป็นสายงานที่ไม่อยู่ในกลุ่มเป้าหมายของการสอบแข่งขันด้วยแบบทดสอบความสามารถทางสมอง

กลุ่มตัวอย่าง

กลุ่มตัวอย่าง ได้แก่ กลุ่มข้าราชการพลเรือนไทยในสายงานระดับแรกบรรจุ ระดับ 3 หรือ 4 ซึ่งยังดำรงตำแหน่งอยู่ในระดับ 3-5 จำนวน 680 คน ซึ่งแบ่งตามขั้นตอนการดำเนินการ ดังนี้

1. ชั้นวิเคราะห์คุณภาพข้อสอบรายข้อ ชั้นวิเคราะห์ความเชื่อมั่นของแบบทดสอบ

และชั้นวิเคราะห์ความเที่ยงตรงตามโครงสร้างของแบบทดสอบ กำหนดกลุ่มตัวอย่างจำนวน 200 คน โดยพิจารณาจากตารางการสุ่มตัวอย่างของยามาเน่ ที่ระดับความเชื่อมั่น 90% และความคลาดเคลื่อน 10% (Yamane, 1960, pp. 1088) ซึ่งกำหนดขนาดของกลุ่มตัวอย่างของประชากรที่ไม่จำกัดจำนวนไว้ 100 คน ประกอบกับการพิจารณาขนาดตัวอย่างที่พอเพียงสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งควรใช้กลุ่มตัวอย่างอย่างน้อย 200 คน จึงจะอยู่ในเกณฑ์ดี

วิธีการสุ่มตัวอย่างใช้การสุ่มตัวอย่างแบบชั้นภูมิ (stratified random sampling) โดยสุ่มสายงาน 40 สายงาน จากกลุ่มสายงาน 8 กลุ่ม โดยคำนวณเทียบสัดส่วนตามจำนวนสายงานที่มีในแต่ละกลุ่มสายงาน (ดังภาคผนวก ข) และกำหนดจำนวนข้าราชการสายงานละ 5 คน รวมเป็น 200 คน ทั้งนี้ ผู้วิจัยจะส่งจดหมายเชิญข้าราชการจากส่วนราชการที่มีสายงานที่สุ่มได้ สายงานละ 7 คน รวม 280 คน เพื่อสำรองในกรณีที่บางส่วนราชการไม่สามารถส่งข้าราชการเข้ารับการทดสอบได้ครบตามจำนวน

2. ชั้นกำหนดเวลาตอบ กำหนดกลุ่มตัวอย่างจำนวน 80 คน

วิธีการสุ่มตัวอย่างใช้การสุ่มตัวอย่างแบบชั้นภูมิ โดยสุ่มสายงานจากกลุ่มสายงาน 8 กลุ่ม กลุ่มละ 2 สายงาน รวม 16 สายงาน (ดังภาคผนวก ข) และกำหนดจำนวนข้าราชการสายงานละ 5 คน ทั้งนี้ ผู้วิจัยจะส่งจดหมายเชิญข้าราชการจาก ส่วนราชการที่มีสายงานที่สุ่มได้ สายงานละ 7 คน รวม 112 คน เพื่อสำรองในกรณีที่บางส่วนราชการไม่สามารถส่งข้าราชการเข้ารับการทดสอบได้ครบตามจำนวน

3. ชั้นวิเคราะห์ความเที่ยงตรงตามสภาพของแบบทดสอบ และชั้นสร้างเกณฑ์ปกติของแบบทดสอบ กำหนดกลุ่มตัวอย่าง จำนวน 400 คน โดยพิจารณาจากตารางการสุ่มตัวอย่างของยามาเน่ที่ระดับความเชื่อมั่น 95% และระดับความคลาดเคลื่อน 5% (Yamane, 1960, pp. 1088) ซึ่งกำหนดขนาดของกลุ่มตัวอย่างของประชากรที่ไม่จำกัดจำนวนไว้ 400 หน่วย

วิธีการสุ่มตัวอย่างใช้การสุ่มตัวอย่างแบบหลายขั้นตอน (multi-stage random sampling) โดยสุ่มกลุ่มสายงาน 4 กลุ่มจากกลุ่มสายงาน 8 กลุ่ม จากนั้นสุ่มสายงานจากกลุ่มสายงาน 4 กลุ่มที่สุ่มได้ กลุ่มละ 3 สายงาน รวม 12 สายงาน (ดังภาคผนวก ข) สายงานละ 35 คน รวมเป็น 420 คน (รวมขนาดตัวอย่าง กลุ่มสายงานละ 105 คน เพื่อให้เพียงพอสำหรับการสร้างเกณฑ์ปกติจำแนกตามกลุ่มสายงาน) ทั้งนี้ ผู้วิจัยจะส่งจดหมายเชิญข้าราชการจากส่วนราชการที่มีสายงานที่สุ่มได้ สายงานละ 40 คน รวม 480 คน เพื่อสำรองในกรณีที่บางส่วนราชการไม่สามารถส่งข้าราชการเข้ารับการทดสอบได้ครบตามจำนวน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ ประกอบด้วย แบบทดสอบความสามารถทั่วไปทางสมองที่ผู้วิจัยสร้างเพื่อนำไปวิเคราะห์คุณภาพรายข้อ และแบบประเมินผลการปฏิบัติงานของข้าราชการพลเรือน ซึ่งมีรายละเอียดดังนี้

1. แบบทดสอบความสามารถทั่วไปทางสมองที่ผู้วิจัยสร้างเพื่อนำไปวิเคราะห์คุณภาพรายข้อ
 - 1.1 ลักษณะของแบบทดสอบ
 - ส่วนที่ 1 เป็นข้อมูลส่วนบุคคล ได้แก่ เพศ อายุ ตำแหน่ง และสังกัด
 - ส่วนที่ 2 เป็นคำชี้แจงในการทดสอบ
 - ส่วนที่ 3 เป็นข้อสอบที่ผู้วิจัยสร้างเพื่อทดสอบความสามารถทั่วไปทางสมอง ด้านการภาษา-การศึกษา (Verbal-educational) จำนวน 120 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก ใช้เวลาในการตอบ 3 ชั่วโมง (ตัวอย่างแสดงในภาคผนวก ค) วัตถุประสงค์ประกอบย่อย 2 องค์กรประกอบ คือ
 1. ความสามารถด้านภาษา จำนวน 60 ข้อ ประกอบด้วย

- การเข้าใจภาษา	20	ข้อ
- อุปมาอุปไมย	20	ข้อ
- การสรุปความ	20	ข้อ
 2. ความสามารถด้านตัวเลข จำนวน 60 ข้อ ประกอบด้วย

- อนุกรมตัวเลข	20	ข้อ
- ความสัมพันธ์เชิงปริมาณ	20	ข้อ
- การแก้โจทย์เลข	20	ข้อ

1.2 เกณฑ์การให้คะแนน การตรวจให้คะแนนแต่ละข้อ ให้ 1 คะแนนเมื่อตอบถูก และให้ 0 คะแนน เมื่อตอบผิด

2. แบบประเมินผลการปฏิบัติงานของข้าราชการพลเรือน สำหรับใช้เป็นเกณฑ์ในการวิเคราะห์ความเที่ยงตรงตามสภาพ (concurrent validity) ของแบบทดสอบความสามารถทั่วไปทางสมองที่ผู้วิจัยสร้าง

2.1 ลักษณะของแบบประเมิน แบบประเมินนี้ เป็นแบบประเมินผลการปฏิบัติงานที่ศูนย์สรรหาและเลือกสรร สำนักงาน ก.พ. เป็นผู้สร้างขึ้นเพื่อใช้ในการศึกษาวิจัย

เกี่ยวกับการสร้างและพัฒนาเครื่องมือประเมินบุคคล โดยใช้ชื่อว่าแบบสำรวจเพื่อการพัฒนา ระบบการสรรหาและเลือกสรรบุคคลเข้ารับราชการ (ตัวอย่างแสดงในภาคผนวก ง) แบ่งเป็น 3 ส่วน คือ

ส่วนที่ 1 เป็นข้อมูลทั่วไปของผู้รับการประเมิน ได้แก่ ชื่อ – นามสกุลของผู้รับการประเมิน ตำแหน่ง ระยะเวลาในการดำรงตำแหน่งในระดับปัจจุบัน ตำแหน่งของผู้ประเมิน และสังกัด

ส่วนที่ 2 เป็นแบบประเมินส่วนผลงานหรืองานที่ได้รับมอบหมาย ประกอบด้วย ข้อความ 6 ข้อ สร้างขึ้นโดยอ้างอิงหลักเกณฑ์การประเมินผลการปฏิบัติงานของข้าราชการตามกฎหมาย ก.พ. ฉบับที่ 10 (พ.ศ. 2538) ซึ่งมุ่งพิจารณาผลงานในด้านปริมาณ คุณภาพ และเวลาที่ใช้ในการปฏิบัติงาน มีค่าความเชื่อมั่นเท่ากับ .83

ส่วนที่ 3 เป็นแบบประเมินส่วนพฤติกรรมการทำงาน ประกอบด้วยข้อความ 53 ข้อ สร้างขึ้นจากการวิเคราะห์งาน มุ่งประเมินพฤติกรรม 10 ด้าน ได้แก่ การสื่อสาร ความละเอียด แม่นยำ การประสานงาน/มนุษยสัมพันธ์ การคิดวิเคราะห์ การแก้ปัญหา ความรับผิดชอบ การมุ่งผลสำเร็จ การให้บริการ ความคิดสร้างสรรค์ และการควบคุมอารมณ์ ข้อความทั้งหมดเป็นข้อความทางบวก มีค่าความเชื่อมั่นอยู่ในช่วง .88-.93 (แสดงดังตารางที่ 3.1)

2.2 เกณฑ์การให้คะแนน

2.2.1 ส่วนที่ 2: ผู้ประเมินพิจารณาระดับผลงานของผู้ใต้บังคับบัญชาโดยเปรียบเทียบกับมาตรฐานการทำงาน ซึ่งแบ่งเป็น 5 ระดับ แล้วทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับผู้ใต้บังคับบัญชามากที่สุด ซึ่งมีเกณฑ์การให้คะแนนดังนี้

ต่ำกว่ามาตรฐานมาก	1
ต่ำกว่ามาตรฐาน	2
เป็นไปตามมาตรฐาน	3
ดีกว่ามาตรฐาน	4
ดีกว่ามาตรฐานมาก	5

2.2.2 ส่วนที่ 3: ผู้ประเมินพิจารณารายการพฤติกรรมในการทำงานที่ละข้อ และประเมินผู้ใต้บังคับบัญชาว่ามีพฤติกรรมเหล่านั้นในระดับใดใน 5 ระดับ โดยทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับผู้ใต้บังคับบัญชามากที่สุด ซึ่งมีเกณฑ์การให้คะแนนดังนี้

น้อยมาก	1
น้อย	2

ปานกลาง	3
มาก	4
มากที่สุด	5
ข้อมูลไม่พอ	ไม่นำมาคิดคะแนน

ทั้งนี้ ผลคะแนนทั้ง 2 ส่วนนี้ จะใช้คะแนนเฉลี่ยของข้อคำถามโดยแบ่งเป็น 3 องค์ประกอบตามผลการวิเคราะห์องค์ประกอบของแบบประเมิน (สำนักงาน ก.พ., 2548, น. 57 - 67) คือ

1. ด้านผลงาน ใช้คะแนนเฉลี่ยจากข้อคำถามด้านผลงาน ความละเอียดแม่นยำ ความรับผิดชอบ และการมุ่งผลสำเร็จ
2. ด้านทักษะความสัมพันธ์ระหว่างบุคคล ใช้คะแนนเฉลี่ยจากข้อคำถามด้านการประสานงาน การให้บริการ และการควบคุมอารมณ์
3. ด้านการคิด ใช้คะแนนเฉลี่ยจากข้อคำถามด้านการสื่อสาร การคิดวิเคราะห์ การแก้ปัญหา และความคิดสร้างสรรค์

ตารางที่ 3.1

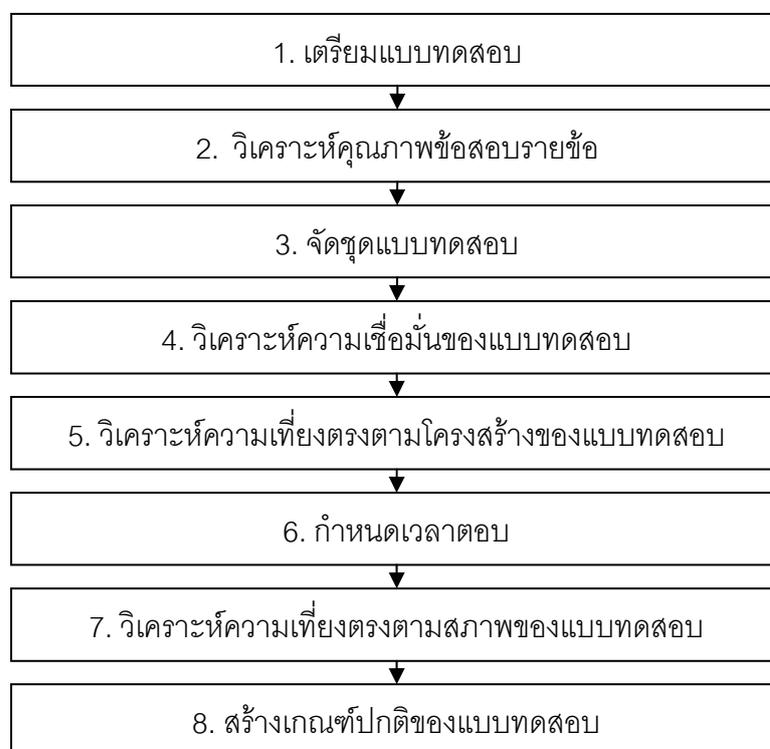
แสดงค่าความเชื่อมั่นของแบบประเมินผลการปฏิบัติงาน

องค์ประกอบ	ค่าความเชื่อมั่น
ผลงาน	.83
การสื่อสาร	.89
ความละเอียดแม่นยำ	.88
การประสานงาน/มนุษยสัมพันธ์	.88
การคิดวิเคราะห์	.89
การแก้ปัญหา	.91
ความรับผิดชอบ	.90
การมุ่งผลสำเร็จ	.90
การให้บริการ	.91
ความคิดสร้างสรรค์	.93
การควบคุมอารมณ์	.92

วิธีการวิจัย

ผู้วิจัยได้กำหนดขั้นตอนการสร้างและพัฒนาแบบทดสอบความสามารถทั่วไปทางสมอง (แสดงในภาพที่ 3.1) ดังนี้

ภาพที่ 3.1
แสดงขั้นตอนการสร้างและพัฒนาแบบทดสอบ



1. ขั้นเตรียมแบบทดสอบ

1.1 ศึกษาปริมาณ ทบพวนทฤษฎี แนวคิดทางสถิติ และงานวิจัยที่เกี่ยวข้อง เพื่อพิจารณาความหมาย องค์ประกอบ และวิธีการสร้างแบบทดสอบความสามารถทั่วไปทางสมอง ซึ่งในการวิจัยนี้ องค์ประกอบของความสามารถทั่วไปทางสมองของแบบทดสอบที่สร้าง ดัดแปลงจากทฤษฎีกลุ่มองค์ประกอบลำดับชั้นของเวอร์นอน (Vernon's hierarchical group factor theory) โดยมุ่งวัดองค์ประกอบของความสามารถทั่วไปทางสมองเฉพาะในลำดับชั้นของกลุ่มองค์ประกอบหลักด้านภาษา-การศึกษา ซึ่งแบ่งกลุ่มองค์ประกอบย่อยเป็น 2 ด้าน ได้แก่ ด้านภาษา และด้านตัวเลข โดยในส่วนขององค์ประกอบเฉพาะที่วัดความองค์ประกอบย่อยด้านภาษานั้น ผู้วิจัย

กำหนดวัดด้วยความสามารถในการทำข้อสอบด้านการเข้าใจภาษา (verbal comprehension) อุปมาอุปไมย (analogies) และการสรุปความ (inference) ส่วนองค์ประกอบย่อยด้านตัวเลข วัดด้วยความสามารถในการทำข้อสอบด้านอนุกรมตัวเลข (number series) ความสัมพันธ์เชิงปริมาณ (quantitative relations) และการแก้โจทย์เลข (mathematics problem solving)

1.2 กำหนดผังโครงสร้างข้อสอบ โดยพิจารณา กำหนดจำนวนข้อสอบให้ใกล้เคียงกับแบบทดสอบ Wonderlic Personnel Test ซึ่งมีจำนวนข้อสอบ 50 ข้อ และพิจารณากำหนดสัดส่วนข้อสอบตามความนิยมในการจัดชุดข้อสอบวัดผลสัมฤทธิ์ ซึ่งมักใช้ข้อสอบที่มีค่าความยากง่ายระหว่าง .20-.40 (25%), .41-.60 (50%) และ .61-.80 (25%) (แสดงในตารางที่ 3.2) จากนั้นสร้างข้อสอบให้มีจำนวนข้ออย่างน้อย 2-3 เท่าของจำนวนข้อที่กำหนดใช้จริง เพื่อสำรองไว้สำหรับการคัดเลือกข้อสอบ (Gregory, 2004, p.126; Kline, 1990, p.88) ทั้งนี้ รูปแบบของข้อสอบ กำหนดเป็นข้อสอบแบบ 4 ตัวเลือก เนื่องจากวัดระดับความสามารถทางสมองได้ดี สามารถอ่านได้ง่าย และเหมาะสำหรับการทดสอบที่จำกัดเวลา (Nunnally and Bernstein, 1994, p. 297)

ตารางที่ 3.2

แสดงผังโครงสร้างข้อสอบ จำแนกตามองค์ประกอบและระดับความยากง่ายของข้อสอบ

องค์ประกอบ	ระดับความยากง่ายของข้อสอบ			
	.20-.40	.41-.60	.61-.80	รวม
1. ความสามารถด้านภาษา	6	12	6	24
1.1 การเข้าใจภาษา	2	4	2	8
1.2 อุปมาอุปไมย	2	4	2	8
1.3 การสรุปความ	2	4	2	8
2. ความสามารถด้านตัวเลข	6	12	6	24
2.1 อนุกรมตัวเลข	2	4	2	8
2.2 ความสัมพันธ์เชิงปริมาณ	2	4	2	8
2.3 การแก้โจทย์เลข	2	4	2	8
รวม	12	24	12	48

1.3 ตรวจสอบและคัดเลือกข้อสอบเพื่อจัดพิมพ์เป็นชุดแบบทดสอบ จำนวน 120 ข้อ และกำหนดเวลาตอบ 3 ชั่วโมง โดยพิจารณาจากข้อแนะนำทั่วไปสำหรับการกำหนดเวลาตอบแบบทดสอบว่า ผู้สอบทั่วไปควรทำข้อสอบแบบหลายตัวเลือกได้ 1 ข้อ ในเวลา 1 นาที และสำหรับข้อที่มีความซับซ้อนหรือต้องอ่านข้อความที่ยาว ควรทำได้ภายในเวลา 2 หรือ 3 นาที (Berry, 2003, p. 207)

1.4 ตรวจสอบความเที่ยงตรงตามเนื้อหา (Content Validity) ของข้อคำถามที่สร้างขึ้น โดยนำข้อคำถามพร้อมทั้งนิยามเชิงทฤษฎี นิยามเชิงปฏิบัติการ และผังโครงสร้างข้อสอบ ให้ผู้เชี่ยวชาญด้านแบบทดสอบความสามารถทางสมองของศูนย์สรรหาและเลือกสรร จำนวน 10 ท่าน พิจารณาดัดสินความสอดคล้องตามเนื้อหาเพื่อนำมาคำนวณค่าอัตราส่วนความเที่ยงตรงตามเนื้อหา (Content Validity Ratio: CVR) ตามวิธีการของลอวชี (Lawshe, 1975, pp. 563-575) และคัดเลือกข้อสอบที่มีค่าความเที่ยงตรงตามเนื้อหาตั้งแต่ .62 ขึ้นไป (แสดงในภาคผนวก จ)

2. ชั้นวิเคราะห์คุณภาพข้อสอบรายข้อ

2.1 นำแบบทดสอบ ไปทดสอบกับกลุ่มตัวอย่างจำนวน 200 คน โดยคำนึงถึงความหลากหลายของสายงาน เพื่อให้แบบทดสอบที่สร้างขึ้นสามารถนำไปใช้ได้ในสายงานที่หลากหลายของข้าราชการพลเรือนระดับแรกบรรจุ และแจ้งให้ผู้สอบทราบถึงความสำคัญของผลการทดสอบซึ่งจะนำไปใช้ในการพัฒนาเกณฑ์สำหรับการคัดเลือกบุคลากรภาครัฐ เพื่อให้ผู้สอบใช้ความสามารถอย่างเต็มที่ในการสอบ

2.2 ตรวจให้คะแนนและวิเคราะห์ข้อสอบรายข้อ โดยคำนวณจากดัชนีบ่งชี้คุณภาพของข้อสอบเพื่อคัดเลือกหรือปรับปรุงคุณภาพข้อสอบ (รัตนาศิริพานิช, 2533; Kline, 1990, p.90) ได้แก่

2.2.1 ค่าความยากง่ายของข้อสอบ (difficulty power) คำนวณจากสัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูกต้องต่อจำนวนผู้ตอบทั้งหมด แล้วคัดเลือกข้อสอบไว้เฉพาะข้อที่มีค่าระหว่าง .20-.80 ซึ่งถือว่าเป็นข้อสอบที่มีความยากง่ายพอเหมาะ

2.2.2 ค่าอำนาจจำแนกของข้อสอบ (discrimination power) คำนวณจากผลต่างระหว่างสัดส่วนของจำนวนคนตอบถูกในกลุ่มของผู้ที่ได้คะแนนสูง เรียกว่า “กลุ่มสูง” ต่อจำนวนคนตอบถูกในกลุ่มผู้ที่ได้คะแนนต่ำ เรียกว่า “กลุ่มต่ำ” ซึ่งผู้วิจัยแบ่งกลุ่มโดยใช้เทคนิค 27% และคัดเลือกเฉพาะข้อสอบที่มีค่าตั้งแต่ .20 ขึ้นไป ซึ่งถือว่าเป็นค่าที่อยู่ในเกณฑ์ใช้ได้

2.2.3 ประสิทธิภาพของตัวลวง คำนวณจากสัดส่วนของจำนวนผู้สอบที่เลือกตัวลวงนั้นๆ ต่อจำนวนผู้สอบทั้งหมด ตัวลวงที่ดีควรมีประสิทธิภาพของตัวลวงตั้งแต่ .05 ขึ้นไป

หากข้อสอบข้อใดมีประสิทธิภาพของตัวลวงและค่าอำนาจจำแนกอยู่ในเกณฑ์ที่กำหนดแล้ว แต่มีประสิทธิภาพของตัวลวงต่ำกว่า .05 ผู้วิจัยจะปรับปรุงตัวลวงในข้อนั้น และให้ผู้เชี่ยวชาญของสำนักงาน ก.พ. เป็นผู้ตรวจสอบความถูกต้อง

3. **ขั้นจัดชุดแบบทดสอบ** คัดเลือกข้อสอบที่มีค่าความยากง่ายตามผังโครงสร้างข้อสอบในตารางที่ 3.1 และมีคุณสมบัติผ่านเกณฑ์บ่งชี้คุณภาพข้อสอบ จำนวน 48 ข้อ มาจัดเรียงแบบชุดเกลียว (spiral omnibus) โดยนำข้อที่ง่ายที่สุดของทุกองค์ประกอบมาจัดเรียงจากข้อที่ง่ายไปสู่ข้อที่ยากกว่า จนครบทุกองค์ประกอบ และเริ่มจัดเรียงข้อที่เหลือวนในลักษณะเดียวกันนี้จนครบทุกข้อ ข้อสอบที่จัดเรียงจะมีลักษณะเป็นขั้นที่แต่ละขั้นประกอบด้วยข้อสอบครบทุกองค์ประกอบ วนเรียงคล้ายเกลียวจากขั้นที่ง่ายที่สุดไปสู่ขั้นที่ยากที่สุด แบบทดสอบ Wonderlic Personnel Test เป็นตัวอย่างที่ดีของแบบทดสอบที่มีการจัดเรียงข้อสอบแบบนี้ ซึ่งมีข้อดีในการช่วยตัดประเด็นปัญหาเกี่ยวกับการจัดลำดับองค์ประกอบในการวัด และช่วยให้สามารถวัดความสามารถของผู้สอบได้ในเวลาที่สั้นกว่าการจัดเรียงข้อสอบแบบแบ่งตอน (section) ซึ่งจัดให้ข้อสอบแต่ละตอนวัดองค์ประกอบเดียวกัน ทั้งนี้เพราะการจัดเรียงข้อสอบแบบแบ่งตอนนั้น ผู้สร้างแบบทดสอบต้องกำหนดเวลาการตอบให้ยาวนานเพียงพอต่อการทำข้อสอบให้ครบทุกตอน จึงจะสามารถวัดความสามารถของผู้สอบได้ครบทุกองค์ประกอบ แต่สำหรับการจัดเรียงข้อสอบแบบชุดเกลียว ผู้สร้างแบบทดสอบสามารถกำหนดเวลาการตอบให้สั้นลงได้โดยไม่จำเป็นต้องให้ผู้สอบตอบแบบทดสอบให้ครบทุกข้อ เนื่องจากภายในเวลาทำตอบที่กำหนดนั้น ผู้สอบมีโอกาสในการตอบข้อสอบครบทุกองค์ประกอบ ซึ่งข้อสอบเหล่านั้นมีการจัดเรียงลำดับตามระดับความยากไว้แล้ว ทั้งนี้ ข้อสอบแต่ละองค์ประกอบในขั้นหนึ่งๆ ไม่จำเป็นต้องกำหนดไว้เพียงข้อเดียว แต่สามารถกำหนดไว้เป็นชุดได้ เช่น ชุดละ 5 ข้อ (Anastasi, 1984, p.272; Anastasi and Urbina, 1997, 300; Berry, 2004, p. 201) ผู้วิจัยจึงกำหนดจำนวนข้อสอบแต่ละองค์ประกอบในขั้นหนึ่งๆ ไว้ชุดละ 2 ข้อ เพื่อให้สอดคล้องกับสัดส่วนจำนวนข้อสอบตามผังโครงสร้างข้อสอบในตารางที่ 3.2 และสอดคล้องกับการวิเคราะห์ความเชื่อมั่นแบบแบ่งครึ่งข้อคือและข้อคู่ในขั้นที่ 4

4. **ขั้นวิเคราะห์ความเชื่อมั่น (reliability) ของแบบทดสอบ** นำคะแนนของข้อสอบที่จัดชุดแบบทดสอบมาวิเคราะห์ค่าความเชื่อมั่นแบบแบ่งครึ่งข้อสอบ (split-half reliability) โดยหาค่าสหสัมพันธ์ระหว่างคะแนนรวมของชุดที่แบ่งครึ่งตามสูตรของสเปียร์แมน-บราวน์ (Spearman-Brown) เนื่องจากแบบทดสอบที่สร้างประกอบด้วยข้อสอบที่มีลักษณะหลากหลาย (heterogeneity of test item) หากวิเคราะห์ความเชื่อมั่นแบบสอดคล้องภายในตามสูตรของคูเดอริชาร์ดสัน ซึ่งเป็นการหาความสัมพันธ์ระหว่างข้อสอบแต่ละข้อกับคะแนนรวม ค่าความเชื่อมั่นที่ได้มักต่ำกว่า

ความเป็นจริง ทั้งนี้ ในการแบ่งครึ่งข้อสอบ ผู้วิจัยใช้การแบ่งครึ่งข้อคือและข้อคู่ เนื่องจากแบบทดสอบที่สร้างขึ้น มีการออกแบบให้จัดเรียงข้อสอบแบบชุดเกลียวโดยจำนวนข้อสอบแต่ละองค์ประกอบในชั้นหนึ่งๆ กำหนดไว้ชุดละ 2 ข้อ การแบ่งครึ่งข้อสอบแบบข้อคือและข้อคู่จึงทำให้ชุดข้อสอบที่แบ่งครึ่งมีองค์ประกอบและความยากง่ายใกล้เคียงกันมากกว่าการแบ่งข้อสอบแบบครึ่งแรกครึ่งหลัง (Anastasi and Urbina, 1997, pp. 95-96)

5. ขั้นวิเคราะห์ความเที่ยงตรงตามโครงสร้าง (construct validity) ของแบบทดสอบ
นำคะแนนของข้อสอบที่จัดชุดแบบทดสอบมาทำการวิเคราะห์องค์ประกอบเชิงยืนยัน(Confirmatory Factor Analysis: CFA) โดยมีสมมุติฐานว่า โมเดลการวัดความสามารถทั่วไปทางสมองด้านภาษา-การศึกษา แบ่งออกเป็น 2 องค์ประกอบ และแต่ละองค์ประกอบแบ่งออกเป็น 3 องค์ประกอบย่อย จึงทำการวิเคราะห์เพื่อยืนยันว่าแบบทดสอบมีโครงสร้างสอดคล้องกับโมเดลที่กำหนดไว้หรือไม่

6. ขั้นกำหนดเวลาการตอบ (time limits)

6.1 นำแบบทดสอบที่สร้างขึ้นจากขั้นตอนที่ 3-5 ไปทดสอบกับกลุ่มตัวอย่างจำนวน 80 คน โดยในการทดสอบครั้งเดียวกันนี้ จะแบ่งเวลาเป็น 6 ช่วง ได้แก่ ช่วงเวลาตอบที่ 15, 20, 25, 30 และ 35 นาที และช่วงที่ไม่จับเวลา (นับต่อจากช่วงเวลาตอบที่ 35 นาที จนทำข้อสอบเสร็จ) แต่ละช่วงเวลาจำแนกด้วยสีของปากกา ได้แก่ สีน้ำเงิน สีดำ สีแดง สีม่วง สีส้ม และสีเขียว ตามลำดับ ซึ่งผู้วิจัยจะแจกให้ผู้สอบเฉพาะการทดสอบในแต่ละช่วงเวลาเท่านั้น โดยเมื่อหมดเวลาในแต่ละช่วง ผู้ดำเนินการสอบจะสั่งให้ผู้สอบหยุดทำตอบ พร้อมทั้งเก็บปากกาสีเดิมและแจกปากกาสีใหม่ให้ และก่อนที่จะเริ่มจับเวลาให้ผู้สอบทำข้อสอบต่อในช่วงต่อไป จะมีการชี้แจงให้ผู้สอบทราบว่าสามารถย้อนกลับไปแก้ไขคำตอบสำหรับข้อที่ผ่านมาแล้วได้โดยใช้ปากกาสีใหม่ที่แจกให้ แต่ไม่ให้ขีดฆ่าหรือลบคำตอบเดิม เพื่อที่ผู้ตรวจจะสามารถสังเกตคำตอบในแต่ละช่วงเวลาได้ ทั้งนี้จะไม่แจ้งให้ผู้สอบทราบว่าหลังจากการทดสอบในช่วงที่จับเวลาแล้ว จะให้ทดสอบต่อโดยไม่จับเวลา เพื่อให้ผู้สอบใช้ความสามารถอย่างเต็มที่ในการตอบข้อสอบในช่วงที่มีการจับเวลา

6.2 ตรวจให้คะแนน และหาคะแนนรวมโดยนับข้อถูกของเวลาตอบแต่ละช่วง
ดังนี้

ช่วงเวลาตอบที่ 15 นาที ให้นับข้อถูกจากปากกาสีน้ำเงิน

ช่วงเวลาตอบที่ 20 นาที ให้นับข้อถูกจากปากกาสีน้ำเงินและสีดำ (ในข้อที่มีการแก้คำตอบ ให้ยึดตามปากกาสีดำเป็นหลัก)

ช่วงเวลาตอบที่ 25 นาที ให้นับข้อถูกจากปากกาสีน้ำเงิน สีดำ และสีแดง (ในข้อที่มีการแก้คำตอบ ให้ยึดตามปากกาสีแดงเป็นหลัก)

ช่วงเวลาตอบที่ 30 นาที ให้นับข้อถูกจากปากกาสีน้ำเงิน สีดำ สีแดง และสีม่วง (ในข้อที่มีการแก้คำตอบ ให้ยึดตามปากกาสีม่วงเป็นหลัก)

ช่วงเวลาตอบที่ 35 นาที ให้นับข้อถูกจากปากกาสีน้ำเงิน สีดำ สีแดง สีม่วง และสีส้ม (ในข้อที่มีการแก้คำตอบ ให้ยึดตามปากกาสีส้มเป็นหลัก)

ช่วงที่ไม่จับเวลา ให้นับข้อถูกจากปากกาสีน้ำเงิน ปากกาสีดำ สีแดง สีม่วง สีส้ม และสีเขียว (ในข้อที่มีการแก้คำตอบ ให้ยึดตามปากกาสีเขียวเป็นหลัก)

6.3 หาค่าสหสัมพันธ์ระหว่างคะแนนรวมของช่วงที่จับเวลาตอบแต่ละช่วงกับช่วงที่ไม่จับเวลา และทำกราฟเปรียบเทียบค่าเบี่ยงเบนมาตรฐานของคะแนนในช่วงที่จับเวลาตอบแต่ละช่วง เพื่อกำหนดเวลาตอบที่เหมาะสม โดยพิจารณาเลือกช่วงเวลาตอบที่คะแนนรวมมีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับคะแนนรวมของการตอบช่วงที่ไม่จับเวลา ซึ่งแสดงให้เห็นว่าการกำหนดความเร็วในการทดสอบตามช่วงเวลาดังกล่าว ไม่เกี่ยวข้องกับความแม่นยำหรือความสามารถของผู้สอบ (Nunnally and Bernstein, 1994, p.348) ประกอบกับการพิจารณาช่วงเวลาตอบที่คะแนนรวมมีค่าเบี่ยงเบนมาตรฐานสูง ซึ่งแสดงให้เห็นว่าในช่วงเวลาตอบดังกล่าว การกระจายคะแนนของผู้สอบมีความแตกต่างกันหรือหลากหลาย (Anastasi and Urbina, 1997, p.104; Nunnally and Bernstein, 1994, p.350)

7. ^๕ ขั้นวิเคราะห์ความเที่ยงตรงตามสภาพ (concurrent validity) ของแบบทดสอบ

7.1 นำแบบทดสอบที่สร้างขึ้นไปดำเนินการทดสอบกับข้าราชการพลเรือนระดับแรกบรรจุ จำนวน 400 คน และส่งแบบประเมินผลการปฏิบัติงานให้กลุ่มผู้บังคับบัญชาระดับต้น ประเมินผลการปฏิบัติงานของข้าราชการกลุ่มดังกล่าว และส่งผลการประเมินกลับมาทางไปรษณีย์โดยตรง

7.2 ตรวจให้คะแนนแบบทดสอบที่สร้างขึ้น และแบบประเมินผลการปฏิบัติงาน

7.3 วิเคราะห์ความเที่ยงตรงตามสภาพของแบบทดสอบ โดยหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันระหว่างคะแนนรวมของแบบทดสอบที่สร้างขึ้นและคะแนนผลการปฏิบัติงานทั้งในภาพรวมและจำแนกตามกลุ่มสายงาน

8. ^๕ ขั้นสร้างเกณฑ์ปกติ (norms) ของแบบทดสอบ นำผลคะแนนรวมจากการทดสอบข้าราชการในขั้นตอนที่ 7 มาแปลงคะแนนเป็นเปอร์เซ็นต์ไทล์รวม และเปรียบเทียบคะแนนเฉลี่ยระหว่างกลุ่มสายงานของข้าราชการ โดยใช้การวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA) เพื่อพิจารณาแปลงคะแนนเป็นเปอร์เซ็นต์ไทล์จำแนกตามกลุ่มสายงานในกรณีที่แต่ละกลุ่มมีคะแนนเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้งนี้ ในส่วนของเพศ และกลุ่มอายุ อาจมีคะแนนเฉลี่ยของ

ความสามารถทั่วไปทางสมองแตกต่างกันด้วย เนื่องจากมีข้อโต้แย้งในประเด็นนี้ค่อนข้างมาก (Ree, Carretta and Steindl, 2001, p. 228) ผู้วิจัยจึงกำหนดที่จะศึกษาเปรียบเทียบคะแนนเฉลี่ยของความสามารถทั่วไปทางสมองระหว่างเพศ และระหว่างกลุ่มอายุเพิ่มเติม โดยใช้สถิติทดสอบที (t-Test) และการวิเคราะห์ความแปรปรวนทางเดียว ตามลำดับ แม้ปัจจุบันภาครัฐจะยังไม่มีแนวโน้มที่จะคัดเลือกข้าราชการโดยพิจารณาจากเพศ และอายุ แต่ผลการศึกษาก็อาจแสดงถึงข้อควรพิจารณาในการใช้แบบทดสอบสำหรับการคัดเลือกต่อไป

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ผู้วิจัยใช้โปรแกรม SPSS for Windows และโปรแกรม LISREL เป็นเครื่องมือในการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังนี้

1. การตรวจสอบการแจกแจงข้อมูลทั่วไป ประกอบด้วยข้อมูล 2 ประเภท คือ
 - 1.1 ข้อมูลเชิงคุณภาพ ได้แก่ เพศ กลุ่มอายุและตำแหน่งงาน เป็นข้อมูลในมาตรฐานบัญญัติ สถิติที่ใช้คือ ค่าความถี่ ค่าร้อยละ
 - 1.2 ข้อมูลเชิงปริมาณ ได้แก่ คะแนนจากแบบทดสอบที่สร้างขึ้น และคะแนนผลการปฏิบัติงาน เป็นข้อมูลในมาตรฐานตรรกาค สถิติที่ใช้คือ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน
2. การตรวจสอบความเที่ยงตรงตามเนื้อหา ใช้การวิเคราะห์ค่าอัตราส่วนความเที่ยงตรงตามเนื้อหาแบบใช้ผู้ตัดสินหลายคนตามวิธีการของลอร์วีซี
3. การตรวจสอบคุณภาพข้อสอบรายข้อ ซึ่งเป็นข้อสอบปรนัยแบบ 4 ตัวเลือก ใช้การวิเคราะห์ค่าความยากง่ายของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และประสิทธิภาพของตัวลอง
4. การตรวจสอบความเชื่อมั่นแบบแบ่งครึ่งข้อคือและข้อคู่ เป็นการวิเคราะห์ข้อมูลในมาตรฐานตรรกาค ใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน-บราวน์ระหว่างคะแนนของข้อคือและข้อคู่
5. การตรวจสอบความเที่ยงตรงตามโครงสร้าง เป็นการวิเคราะห์ข้อในมูลอันตรภาค ใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน
6. การตรวจสอบความสัมพันธ์ระหว่างคะแนนที่ได้จากการตอบแบบทดสอบในช่วงที่มีการจับเวลาและช่วงที่ไม่จับเวลา เป็นการวิเคราะห์ข้อมูลในมาตรฐานตรรกาค ใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เพื่อหาช่วงของการจับเวลาที่คะแนนจากการทดสอบมีความสัมพันธ์กับการทดสอบโดยไม่จับเวลา

7. การตรวจสอบความเที่ยงตรงตามสภาพในการทำนายผลการปฏิบัติงาน เป็นการวิเคราะห์ข้อมูลในมาตรฐานระหว่างคะแนนจากแบบทดสอบที่สร้างขึ้น และคะแนนผลการปฏิบัติงานที่ประเมินโดยหัวหน้างาน ใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เพื่อตรวจสอบว่าแบบทดสอบที่สร้างขึ้น สามารถใช้ทำนาย ผลการปฏิบัติงานของข้าราชการในสภาพปัจจุบันได้ดีเพียงใด

8. การเปรียบเทียบคะแนนเฉลี่ยของแบบทดสอบที่สร้างขึ้น ระหว่างเพศของกลุ่มข้าราชการ ใช้สถิติทดสอบที และการเปรียบเทียบคะแนนเฉลี่ยของแบบทดสอบที่สร้างขึ้น ระหว่างกลุ่มอายุ และกลุ่มสายงานของข้าราชการ ใช้การวิเคราะห์ความแปรปรวนทางเดียว

9. การสร้างเกณฑ์ปกติของแบบทดสอบที่สร้างขึ้น เป็นการจัดตำแหน่งคะแนนของผู้สอบให้อยู่ในรูปของเปอร์เซ็นต์ไทล์ เพื่อใช้เป็นเกณฑ์ในการเปรียบเทียบระดับความสามารถของผู้สอบ

ตารางที่ 3.3

แสดงวัตถุประสงค์ในการวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

วัตถุประสงค์	การวิเคราะห์ทางสถิติ
1. ตรวจสอบการแจกแจงข้อมูลทั่วไป	ความถี่ ร้อยละ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน
2. ตรวจสอบคุณภาพข้อสอบรายข้อ	ค่าความยากง่ายของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ ประสิทธิภาพของตัวลอง
3. ตรวจสอบความเชื่อมั่นแบบแบ่งครึ่ง	ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน-บราวน์
4. ตรวจสอบความเที่ยงตรงตามโครงสร้าง	การวิเคราะห์องค์ประกอบเชิงยืนยัน
5. ตรวจสอบความสัมพันธ์ระหว่างคะแนนที่ได้จากการทดสอบโดยจับเวลาและไม่จับเวลา	ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
6. เปรียบเทียบการกระจายของคะแนนที่ได้จากการทดสอบโดยจับเวลา	ค่าเบี่ยงเบนมาตรฐาน
7. เปรียบเทียบคะแนนเฉลี่ยของแบบทดสอบที่สร้างขึ้น ระหว่างเพศของข้าราชการ	สถิติทดสอบที

ตารางที่ 3.3 (ต่อ)

แสดงวัตถุประสงค์ในการวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

วัตถุประสงค์	การวิเคราะห์ทางสถิติ
8. เปรียบเทียบคะแนนเฉลี่ยของแบบทดสอบ ที่สร้างขึ้น ระหว่างกลุ่มอายุ และกลุ่มสาย งานของข้าราชการ	การวิเคราะห์ความแปรปรวนทางเดียว
9. สร้างเกณฑ์ปกติ	เปอร์เซ็นต์ไทล์