

การตรวจสอบเครื่องมือวัดการรู้คิดแบบ Recall ของผู้สูงอายุไทยจากโครงการสำรวจด้านสุขภาพ การสูงอายุ และการเกษียณในประเทศไทย (HART)

จิตภา วรภักดี¹, รมิดา ศรีเหรา² และ พาชิตชนันต์ ศิริพานิช³

Received: 5 May 2025; Revised: 22 July 2025; Accepted: 4 August 2025

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อตรวจสอบประสิทธิผลของชุดคำ 4 ชุด (ชุดละ 10 คำ) ของแบบทดสอบ (Word Recall) ซึ่งเป็นเครื่องมือที่มีประสิทธิภาพสำหรับวัดการรู้คิด (Cognition) โดยใช้ข้อมูลจากการสำรวจ Panel ผู้สูงอายุไทย ในรอบที่ 4 ซึ่งดำเนินการในปี พ.ศ. 2565 – 2566 ของโครงการสำรวจด้านสุขภาพ การสูงอายุ และการเกษียณในประเทศไทย (HART) ในการวิเคราะห์ได้พิจารณาประสิทธิผลจาก ระดับความยาก (Difficulty Level) อำนาจการจำแนก (Discriminating Power) และความเชื่อถือได้ (Reliability) โดยใช้สถิติเชิงพรรณนาและเปรียบเทียบระดับความยาก อำนาจการจำแนก ของชุดคำด้วยการวิเคราะห์ความแปรปรวนทางเดียว (One-Way ANOVA)

ผลการศึกษาพบว่า ชุดคำทั้ง 4 ชุดประกอบด้วยคำที่มีระดับยากปานกลางคิดเป็นร้อยละ 80 และชุดคำที่ 1 ประกอบด้วยคำที่มีอำนาจการจำแนกดีมากคิดเป็นร้อยละ 80 ในขณะที่ชุดคำอื่นๆ ประกอบด้วยคำที่มีอำนาจจำแนกดี คิดเป็นร้อยละ 60 นอกจากนี้ยังพบว่า ความแตกต่างของค่าเฉลี่ยระดับความยากและอำนาจการจำแนกของชุดคำทั้ง 4 ชุดไม่มีนัยสำคัญทางสถิติ ในส่วนความเชื่อถือได้ พบว่าชุดคำทั้ง 4 ชุดควรได้รับการปรับปรุงในส่วนนี้

คำสำคัญ: การรู้คิด, ผู้สูงอายุ, ระดับความยาก, อำนาจการจำแนก, ความเชื่อถือได้, โครงการสำรวจด้านสุขภาพ การสูงอายุและการเกษียณในประเทศไทย (HART)

Corresponding Author Email: 6510413001@stu.nida.ac.th

¹ นักศึกษาหลักสูตรมหาบัณฑิตสาขาสถิติประยุกต์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

² ผู้ช่วยศาสตราจารย์ และ นักวิจัยอิสระ

³ รองศาสตราจารย์ หลักสูตรมหาบัณฑิตสาขาสถิติประยุกต์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

An Investigation of the Word Recall: Cognitive Assessment in Aged Person from The Health, Aging and Retirement in Thailand (HART) Project

Chidapha Worapukdee¹, Ramidha Srihera² and Pachitjanut Siripanich³

Received: 5 May 2025; Revised: 22 July 2025; Accepted: 4 August 2025

Abstract

This study aims to evaluate the efficacy of the ten-word recall test which is an effective tool in cognition assessment. Data from the Survey from HART project, Panel 2022-2023, Wave 4 have been used for in-depth analysis. Our analysis first focuses on understanding difficulty level, discriminating power and reliability of the four sets of the ten-word recall test by utilizing descriptive statistics. We then compare difficulty level and discriminating power among the four sets of the ten-word recall test by one-way ANOVA.

Our findings reveal that all four sets can be classified as moderate with regards to their difficulty level with eight out of ten words. Moreover, only the first set has very good discriminating power with eight out of ten words whereas the other three sets also have very good discriminating power with six out of ten words. Moreover, both discriminating power and difficulty level have a statistically nonsignificant difference means among the score from four sets of ten-word recall test. Besides, improving reliability is necessary.

Keywords: Cognition, Aged Person, Difficulty Level, Discriminating Power, Reliability, Aging and Retirement in Thailand (HART) Project

Corresponding Author Email: 6510413001@stu.nida.ac.th

¹ Graduate Student, Applied Statistics, Graduate School of Applied Statistics, National Institute of Development Administration

² Assistant Professor and Researcher

³ Associate Professor, Applied Statistics, Graduate School of Applied Statistics, National Institute of Development Administration

1. ความสำคัญของปัญหา

ประเทศไทยเป็นสังคมผู้สูงอายุโดยสมบูรณ์ (Complete Aged Society) มาตั้งแต่ปี พ.ศ. 2565 โดยผู้สูงอายุ 60 ปีขึ้นไปทั่วประเทศมีจำนวน 13,444,127 คน คิดเป็นร้อยละ 20.69 ของประชากรทั้งหมด 64,973,186 คน (กรมกิจการผู้สูงอายุ, 2567) ด้วยโครงสร้างสังคมสูงวัยแบบสมบูรณ์ ทำให้ประเทศไทยต้องเตรียมความพร้อมในการบริหารจัดการเพื่อรับมือรอบด้านไม่ว่าจะเป็นด้านเศรษฐกิจ ด้านสังคม ด้านการแพทย์และสาธารณสุข ฯลฯ

ในทางการแพทย์และสาธารณสุข ปัญหาภาวะสมองเสื่อมของผู้สูงอายุไทยเป็นปัญหาด้านสุขภาพอย่างหนึ่งที่พบมากและมีแนวโน้มเพิ่มขึ้นทุก ๆ ปี นับเป็นอาการป่วยเรื้อรังที่ส่งผลให้ประสิทธิภาพการทำงานของสมองลดลง และสูญเสียความสามารถทางสมอง เช่น ความจำ การรับรู้ ความเข้าใจ การใช้ภาษา ทิศทาง การใช้เหตุผลและการแก้ปัญหา ตลอดจนการเปลี่ยนแปลงในพฤติกรรมและบุคลิกภาพ อาการดังกล่าวมีผลกระทบต่อการทำงานของประจำวันและการเข้าสังคม เป็นภาวะที่ไม่สามารถรักษาให้หายขาดได้ และการดำเนินโรคจะรุนแรงมากขึ้นเมื่อเวลาผ่านไป (Alzheimer's Association, 2017) ในปัจจุบันตัวเลขผู้สูงอายุที่มีภาวะสมองเสื่อมทั่วโลกมีถึง 55 ล้านคน และคาดว่าจะเพิ่มขึ้นประมาณ 10 ล้านคนในทุก ๆ ปี นอกจากนี้ยังพบว่ามีร้อยละ 60 – 70 ของผู้ป่วยโรคอัลไซเมอร์มีสาเหตุมาจากภาวะสมองเสื่อม (World Health Organization, 2023)

สถิติผู้สูงอายุไทยที่มีภาวะสมองเสื่อมปี พ.ศ. 2565 มีประมาณ 770,000 คน คิดเป็นร้อยละ 6 ของจำนวนผู้สูงอายุรวมทั้งประเทศ โดยคาดว่าผู้ที่มีภาวะสมองเสื่อมจะเพิ่มขึ้นเฉลี่ยทุก ๆ ปี ๆ ละประมาณ 100,000 คน (กรมการแพทย์, 2566) ภาวะสมองเสื่อมไม่เพียงแต่มีผลกระทบอย่างมากกับร่างกายดังที่กล่าวข้างต้น แต่ยังมีผลทางด้านจิตใจ ทั้งของผู้ป่วยและผู้ดูแล อย่างไรก็ตามแม้ว่าภาวะสมองเสื่อมจะไม่สามารถรักษาให้หายขาดได้ แต่หากผู้ป่วยได้รับการทดสอบภาวะสมองเสื่อมตั้งแต่เนิ่น ๆ จะช่วยชะลอการเกิดโรคอัลไซเมอร์ได้ ดังนั้นการคัดกรองภาวะสมองเสื่อมในระยะแรกจึงมีความสำคัญและเป็นความจำเป็นอย่างมาก (Belleville et al., 2017)

จากหลักฐานทางการแพทย์พบว่าหนึ่งในลักษณะเด่นของผู้ป่วยอัลไซเมอร์เนื่องมาจากภาวะสมองเสื่อมจะมีความบกพร่องของความจำอาศัยเหตุการณ์ (Episodic Memory) ซึ่งเป็นความจำรวมประสบการณ์ต่าง ๆ ของตัวเองในอดีต (Episodic Memory, 2024) ด้วยเหตุนี้การทดสอบการทำงานของความจำอาศัยเหตุการณ์ได้อย่างถูกต้องจึงเป็นกุญแจสำคัญดอกหนึ่งที่จะช่วยตรวจสอบภาวะสมองเสื่อมซึ่งมีความน่าจะเป็นสูงที่จะเกิดกับผู้สูงอายุ

การวัดการรู้คิด (Cognition) มีความเกี่ยวข้องกับความจำอาศัยเหตุการณ์ของผู้สูงอายุ เป็นเครื่องมือที่ยอมรับอย่างกว้างขวางทั่วโลกที่สามารถช่วยระบุความเสื่อมถอยของการรับรู้ในผู้สูงอายุได้ระดับหนึ่ง และยังช่วยติดตามการเปลี่ยนแปลงของการรับรู้ได้ค่อนข้างชัดเจน (Smith et al., 2018) การวัดการรู้คิดสามารถให้ข้อมูลเชิงลึกเกี่ยวกับความสามารถด้านการรับรู้และเพื่อเป็นแนวทางการตัดสินใจเลือกแผนในการดูแลผู้สูงอายุ ตั้งแต่เริ่มแรกอาจช่วยป้องกันหรือชะลอปรากฏการณ์ความเสื่อมถอยของสมองได้ (Brown, 2019)

โครงการสำรวจด้านสุขภาพ การสูงอายุ และการเกษียณในประเทศไทย (Health, Aging and Retirement in Thailand, HART) ซึ่งเป็นโครงการที่อยู่ภายใต้การดำเนินงานของศูนย์วิจัยสังคมสูงอายุ สถาบันบัณฑิตพัฒนบริหารศาสตร์ (<https://hart.nida.ac.th/>) ดำเนินการสำรวจแผง (Panel Survey) โดยมีกลุ่มเป้าหมายเป็นผู้ที่มีอายุตั้งแต่ 45 ปีขึ้นไปจาก 5 ภูมิภาครวมถึงกรุงเทพมหานคร ดำเนินการสำรวจทุก ๆ 2 ปีโดยประมาณ และเริ่มการสำรวจรอบที่ 1 (Wave 1) ปี พ.ศ. 2557 โครงการ HART ตระหนักและเล็งเห็นความสำคัญของการศึกษาเรื่องดังกล่าว จึงนำแบบทดสอบการวัดการรู้คิด (Cognition) บางอย่างมาเป็นส่วนหนึ่งของแบบสอบถามเพื่อเก็บรวบรวมข้อมูลของกลุ่มเป้าหมาย (ผู้ที่มีอายุตั้งแต่ 45 ปีขึ้นไป) ให้นักวิจัยที่สนใจศึกษาเกี่ยวกับการรู้คิดสามารถนำข้อมูลเหล่านี้ไปวิจัยต่อยอดได้ ส่วนหนึ่งของการแบบทดสอบการวัดการรู้คิดที่โครงการ HART นำมาใช้ในแบบสอบถามคือการรู้คิดแบบ "Recall" ซึ่งประกอบด้วยชุดคำ 4 ชุด ๆ ละ 10 คำ โดยเริ่มดำเนินการครั้งแรกในการสำรวจรอบที่ 2 (Wave 2) ปี พ.ศ. 2559 และดำเนินการต่อเนื่องมาจนถึงรอบที่ 4 นับเป็นระยะเวลาเกือบ 10 ปี ดังนั้น เพื่อให้แน่ใจว่าข้อมูลในส่วนนี้ของโครงการ HART สามารถนำไปใช้ให้เกิดผลสัมฤทธิ์ที่ถูกต้องแม่นยำและเชื่อถือได้จึงสมควรที่จะต้องตรวจสอบเครื่องมือ

การวัดการรู้คิดแบบ Recall ของโครงการ HART ทั้ง 4 ชุดว่ามีประสิทธิภาพเท่า ๆ กันหรือไม่และอยู่ในระดับใด เพื่อนำผลจากการศึกษาครั้งนี้ไปเป็นแนวทางในการปรับปรุงแบบทดสอบการวัดการรู้คิดให้มีประสิทธิภาพดียิ่งขึ้น

2. วัตถุประสงค์

1. เพื่อประเมินประสิทธิภาพของแบบทดสอบการวัดการรู้คิดแบบ Recall ที่ใช้ในโครงการ HART โดยพิจารณาจากระดับความยาก อำนาจการจำแนก ความเชื่อถือได้ของแบบทดสอบทั้ง 4 ชุดซึ่งแต่ละชุดประกอบด้วยคำ 10 คำ
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบทดสอบการวัดการรู้คิดแบบ Recall ที่ใช้ในโครงการ HART ว่ามีระดับความยาก ระดับอำนาจการจำแนก เหมือนหรือแตกต่างกัน

3. ทบทวนวรรณกรรม

3.1 ฟังก์ชันการรู้คิด (Cognitive Function)

ฟังก์ชันการรู้คิดคือกระบวนการที่สมองใช้ในการประมวลผลข้อมูลและตัดสินใจสิ่งต่าง ๆ ซึ่งถูกคิดค้นโดยนักจิตวิทยาชื่อ Carl Jung (Jung, 1971) การทำงานเหล่านี้มีทั้งหมด 8 แบบ แบ่งตามทิศทางการทำงาน ได้แก่ (1) Introversion (2) Extraversion (3) Sensing (4) Intuition (5) Thinking (6) Feeling (7) Judging และ (8) Perceiving โดยแต่ละฟังก์ชันทำงานร่วมกันในรูปแบบ Type Dynamic หรือพลวัตของทฤษฎี Myers-Briggs Type Indicator (MBTI) (Myers & Myers, 2010) การทำความเข้าใจ ฟังก์ชันการรู้คิดมีความสำคัญเนื่องจากช่วยให้บุคคลรู้จักตนเองมากขึ้น โดยเฉพาะเมื่อต้องเผชิญกับสภาวะเครียดและกับดักของบุคลิกภาพที่มักพบและส่งเสริมการพัฒนาดตนเองและการโค้ชบุคลิกภาพ โดยสามารถปรับตัวและพัฒนาฟังก์ชันที่ไม่ถนัดเพื่อการอยู่ร่วมกับผู้อื่นได้ดีขึ้น นอกจากนี้ ฟังก์ชันการรู้คิดยังช่วยในการวิเคราะห์ความแตกต่างของบุคลิกภาพทั้ง 16 แบบใน MBTI ที่คำอธิบายจากแหล่งต่าง ๆ อาจไม่ตรงกับลักษณะของบุคคลนั้นในบางช่วงเวลา การเข้าใจ Cognitive Function ช่วยให้บุคคลสามารถพัฒนาดตนเองและปรับปรุงบุคลิกภาพได้อย่างมีประสิทธิภาพ (Hunziker & Hass, 2014)

เครื่องมือหนึ่งที่ได้รับการยอมรับมากที่สุดและถูกนำไปใช้ทั้งงานวิจัยทางคลินิกตลอดจนงานวิจัยด้านอื่นๆ สำหรับวัดการทำงานที่เกี่ยวข้องกับความจำอาศัยเหตุการณ์ (Episodic Memory) คือ การทดสอบการจดจำคำ (Word List Memory Test, WLMT) ของ CERAD (The Consortium to Establish a Registry for Alzheimer 's Disease) โดย CERAD ได้พัฒนา WLMT ขึ้นเฉพาะสำหรับผู้ป่วยที่มีแนวโน้มจะเป็นโรคอัลไซเมอร์ ปัจจุบันแบบทดสอบนี้ได้รับการแปลเป็นภาษาต่างประเทศอย่างน้อย 24 ภาษาทั่วโลกรวมถึงภาษาไทย (Wagel et.al., 2023) โดยปรกติ WLMT จะประกอบด้วยส่วนประกอบหลัก 8 ส่วน ซึ่งผู้ใช้งานสามารถเลือกแบบทดสอบให้เหมาะสมตามวัตถุประสงค์ของงาน โดยงานวิจัยนี้เลือกศึกษาในส่วนที่ 6 Word List Recall ซึ่งเป็นการกล่าวคำนามที่รับฟังไป (Free Recall of the Nouns) ก่อนหน้านั้นทั้งโดยทันที (Immediate Recall) และเว้นระยะเวลา (Delayed Recall) (CERAD, 2025)

3.2 ดัชนีสำหรับประเมินประสิทธิภาพของเครื่องมือ

เนื่องจากแบบทดสอบการรู้คิด (Cognition) แบบ Recall ที่ต้องการประเมินประสิทธิภาพมีลักษณะคล้ายข้อสอบแบบปรนัย (Multiple Choice Question, MCQ) ที่มีข้อเลือก (Choice) 2 ข้อคือ “จำได้” กับ “จำไม่ได้” นั่นคือ ในแต่ละชุดคำที่ประกอบด้วยคำศัพท์ 10 คำเสมือนให้เป็นข้อสอบ 10 ข้อ ถ้าผู้ตอบจำคำใดได้หมายความว่าตอบข้อสอบนั้นถูกและจะได้คะแนน 1 คะแนน แต่ถ้าจำคำไม่ได้หมายความว่าตอบข้อสอบนั้นไม่ถูกต้อง จะได้ 0 คะแนน ดังนั้นในการประเมินประสิทธิภาพของแบบทดสอบการรู้คิดแบบ Recall นี้จึงอาศัยเครื่องมือแบบเดียวกับที่ใช้ในการวิเคราะห์ข้อสอบ

ในด้านการศึกษา (Education) การวัดผลและการวิเคราะห์ข้อสอบที่ใช้ในการวัดผลสัมฤทธิ์ถือเป็นส่วนสำคัญ (Popham, 2002 และ Trice, 2000) โดยคุณภาพของแบบทดสอบหรือข้อสอบทั้งฉบับขึ้นอยู่กับคุณภาพของข้อถามแต่

ละข้อ (Freeman, 1962 และ Sharma, 2000) การวิเคราะห์ข้อสอบทางการศึกษาจึงเป็นกระบวนการทดสอบคุณภาพทั้งของข้อถามแต่ละข้อและของแบบทดสอบโดยรวม (Singh, Gupta & Singh, 2009) และนับว่ามีคุณค่าและค่อนข้างง่ายที่ควรทำหลังการสอบเพื่อชี้ให้เห็น Validity และ Reliability ของข้อสอบที่ใช้วัดผลสัมฤทธิ์ของการศึกษา (Considine, Botti & Thomas, 2005) การวิเคราะห์ข้อสอบนั้นช่วยให้สามารถสังเกตลักษณะของข้อถามแต่ละข้อและทำให้สามารถปรับปรุงคุณภาพของแบบทดสอบ/ข้อสอบได้ (Gronlund, 1993) เครื่องมือที่ใช้ในการวิเคราะห์ข้อสอบจะช่วยให้ระบุว่าข้อถามข้อใดยากเกินไปหรือง่ายเกินไป และข้อถามใดที่ไม่สามารถแบ่งแยก (Discriminate) ระหว่างผู้ที่ “รู้” และผู้ที่ “ไม่รู้” ได้ (Lange & Mehrens, 1967) ผลการวิเคราะห์ข้อสอบจะให้ข้อเสนอแนะแก่ผู้สอนในการปรับปรุง และเปลี่ยนคำถาม/ข้อถาม MCQ ให้เหมาะสมกับการสอบ คำถาม/ข้อถาม MCQ บางข้ออาจถูกแก้ไขหรือลบออกได้ (Zubairi & Kassim, 2006 และ Sim & Rasiah, 2006)

เนื่องจากแบบทดสอบการรู้คิดแบบ Recall มีลักษณะคล้ายข้อสอบ MCQ ด้วยเหตุนี้ผู้วิจัยจึงนำหลักของการวิเคราะห์ข้อสอบมาประยุกต์ใช้ โดยพิจารณาว่าค่าต่าง ๆ ในแต่ละชุดค่าเป็นค่าที่เหมาะสมไม่ง่ายและไม่ยากหรือซับซ้อนเกินไป ตลอดจนเป็นค่าที่สามารถแบ่งแยกตัวอย่างตามศักยภาพด้านการรู้คิดได้ ดังนั้นผู้วิจัยจึงพิจารณาใช้ระดับความยาก (Difficulty Level) อำนาจการจำแนก (Discriminating Power) และความเชื่อถือได้ (Reliability) เป็นดัชนีสำหรับแบบทดสอบการรู้คิดแบบ Recall ซึ่งมีรายละเอียดดังต่อไปนี้

- **ระดับความยาก (Difficulty Level)**

ระดับความยากของคำพิจารณาจากสัดส่วนจำนวนผู้สูงอายุที่สามารถจดจำคำนั้นได้ (กล่าวทวนออกมาได้ถูกต้อง) เทียบกับจำนวนผู้สูงอายุทั้งหมดที่ทดสอบการจำได้ของคำนั้น หากพบว่าจำนวนผู้สูงอายุที่จำคำนั้นได้มีค่าน้อย หมายความว่าคำนั้นเป็นคำที่ยาก โดยสูตรและเกณฑ์ในการวัดระดับความยากของคำเป็นดังต่อไปนี้ (Gronlund, 1993)

ระดับความยากของคำที่ i ในชุดค่าที่ k

$$DL_{ik} = \frac{\text{จำนวนผู้ที่ตอบแบบทดสอบการรู้คิดแบบ Recall ที่จำคำที่ } i \text{ ในชุดที่ } k \text{ ได้}}{\text{จำนวนผู้ที่ตอบแบบทดสอบการรู้คิดแบบ Recall ด้วยชุดค่าที่ } k \text{ ทั้งหมด}}$$

โดย $i = 1, 2, \dots, 10$ และ $k = 1, 2, 3, 4$

ตารางที่ 1 เกณฑ์ของระดับความยาก ความหมาย และคำแนะนำเป็นดังนี้

ค่าระดับความยาก (DL)	ความหมาย	คำแนะนำ
0.00 – 0.20	ยากมาก	ไม่ควรนำไปใช้
0.21 – 0.30	ยาก	ควรพิจารณาปรับปรุง
0.31 – 0.70	ปานกลาง	นำไปใช้ได้
0.71 – 0.89	ง่าย	ควรพิจารณาปรับปรุง
0.90 – 1.00	ง่ายมาก	ไม่ควรนำไปใช้

ที่มา: Basuki & Hariyanto (2015, cited in Friatma & Anhar, 2019, p.4)

- **อำนาจการจำแนก (Discriminating Power)**

อำนาจการจำแนกพิจารณาจากสัดส่วนของผลต่างจำนวนผู้สูงอายุที่ตอบถูกต้องในกลุ่มบนและกลุ่มล่าง เทียบกับครึ่งหนึ่งของจำนวนผู้สูงอายุทั้งหมดที่อยู่ในกลุ่มบนและกลุ่มล่าง หากพบว่าค่า ๆ นั้นมีค่าอำนาจการจำแนกสูง หมายความว่าคำนั้นเป็นคำที่ดีมาก เนื่องจากสามารถใช้เป็นคำที่แบ่งแยกได้ชัดเจนระหว่างผู้ที่จำคำได้และจำคำไม่ได้ โดยสูตรและเกณฑ์ในการวัดอำนาจการจำแนกเป็นดังต่อไปนี้ (Gronlund, 1993: 103)

อำนาจการจำแนกของคำที่ i ในชุดคำที่ k

$$DP_{ik} = \frac{R_U - R_L}{\left(\frac{1}{2}\right) T}$$

โดย DP_{ik} = อำนาจการจำแนกของคำที่ i ในชุดคำที่ k ($i = 1, 2, \dots, 10$ และ $k = 1, 2, 3, 4$)

R_U = จำนวนผู้ที่จำคำที่ i ในชุดคำที่ k ได้ที่อยู่ในกลุ่มบน (กลุ่มบน ได้แก่ ผู้ที่มีคะแนนการจำคำได้อยู่ในร้อยละ 27 คนแรกของผู้ที่มิฉะนั้นสูงสุด)

R_L = จำนวนผู้ที่จำคำที่ i ในชุดคำที่ k ได้ที่อยู่ในกลุ่มล่าง (กลุ่มล่าง ได้แก่ ผู้ที่มีคะแนนการจำคำได้อยู่ในร้อยละ 27 คนแรกของผู้ที่มิฉะนั้นต่ำสุด)

T = จำนวนผู้ที่จำคำที่ i ในชุดคำที่ k ได้ที่อยู่ในกลุ่มบนและกลุ่มล่างรวมกัน = $R_U + R_L$

ตารางที่ 2 เกณฑ์ของอำนาจการจำแนก ความหมาย และคำแนะนำ เป็นดังนี้

ค่าอำนาจการจำแนก (DP)	ความหมาย	คำแนะนำ
> 0.39	ดีมาก	ควรรักษาไว้
$0.30 - 0.39$	ดี	ควรปรับปรุงเมื่อมีโอกาส
$0.20 - 0.29$	ปานกลาง	ควรทบทวนการนำไปใช้
$0.00 - 0.19$	ค่อนข้างไม่ดี	ไม่ควรนำไปใช้ หรือต้องทบทวนการนำไปใช้
$< - 0.01$	ไม่ดี	ต้องไม่นำไปใช้

หมายเหตุ: ปรับปรุงแก้ไขจาก (Backhoff et al.,2000)

- ความเชื่อถือได้ (Reliability)

การวัดค่าความเชื่อถือได้จะพิจารณาจากความสม่ำเสมอภายใน (Internal Consistency) ของแบบทดสอบที่มีการให้คะแนนแบบ Dichotomous เช่น มีคำตอบเพียงสองแบบ เช่น ถูกหรือผิด ใช่หรือไม่ใช่ และในที่นี้คือจำคำที่ i ได้ (กล่าวคำที่ i ออกมาถูกต้อง) หรือจำไม่ได้ (ไม่ได้กล่าวคำที่ i ออกมา) โดยสูตรและเกณฑ์ในการวัดความเชื่อถือได้ของแบบทดสอบเป็นไปตามสูตร Kuder-Richardson Formula 20 (KR-20)

ความเชื่อถือได้ของชุดคำที่ k

$$R_k = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K p_{ik} (1 - p_{ik})}{s_k^2} \right)$$

โดย K = จำนวนคำทั้งหมดในชุดคำที่ k ($k = 1, 2, 3, 4$) และในที่นี้ $K = 10$

p_{ik} = สัดส่วนของผู้ตอบแบบทดสอบที่จำคำที่ i ได้ ($i = 1, \dots, K$) ในชุดคำที่ k

s_k^2 = ค่าความแปรปรวนของคะแนนในชุดคำที่ k

Σ = สัญลักษณ์แทนการบวก

ตารางที่ 3 เกณฑ์ความเชื่อถือได้ และ ความหมาย เป็นดังนี้

ค่าความเชื่อถือได้	ความหมาย
0.90-1.00	ดีมาก
0.80-0.89	ดี
0.70-0.79	ปานกลาง
0.60-0.69	ไม่แน่ใจใน
0.50-0.59	ไม่ดี
< 0.50	ไม่สามารถยอมรับได้

ที่มา: Kuder & Richardson, 1937

4. วิธีดำเนินการศึกษา

4.1 ข้อมูล แหล่งข้อมูล และตัวแปรของการศึกษา

การศึกษานี้ใช้ข้อมูล Part Cc เรื่องการรู้คิด (Cognition) แบบ Immediate Recall จากโครงการสำรวจด้านสุขภาพ การสูงอายุ และการเกษียณในประเทศไทย (Health, Aging, and Retirement in Thailand, HART) รอบ (Wave) ที่ 4 ซึ่งดำเนินการในปี พ.ศ. 2565 – 2566 (<https://hart.nida.ac.th/>) โดยเป็นข้อมูลของผู้มีอายุตั้งแต่ 45 ปีขึ้นไปจำนวน 4,312 คน

การศึกษากการรู้คิด (Cognition) แบบ Recall ของ HART ประกอบด้วยชุดคำ 4 ชุด แต่ละชุดประกอบด้วยคำต่าง ๆ 10 คำและดำเนินการทดสอบโดยพนักงานสัมภาษณ์เอย่คำทั้ง 10 คำในชุดคำที่กำหนดให้ผู้ตอบแบบสอบถามฟัง 1 รอบแล้วให้ผู้ตอบแบบสอบถามเอย่คำดังกล่าวโดยไม่ต้องเรียงลำดับ คำที่เอย่ออกมาถูกต้องตรงกับคำในชุดคำนั้นเรียกว่า “จำได้” และได้คะแนน 1 คะแนนส่วนคำที่จำไม่ได้หรือไม่เอย่ออกมาจะได้ 0 คะแนน ชุดคำที่กำหนดให้ผู้ตอบแบบสอบถามแต่ละคนได้จากการสุ่มชุดคำในการสำรวจรอบ (Wave) ที่ 2 หรือรอบแรกที่เข้ามาเป็นหน่วยตัวอย่าง (สำหรับผู้ที่เข้ามาหลังรอบที่ 2) และจะเป็นชุดคำที่กำหนดให้ผู้ตอบแบบสอบถามในรอบของการสำรวจต่อ ๆ มาทุกรอบ โดยคำและชุดคำทั้งหมดแสดงไว้ในตารางที่ 4

ตารางที่ 4 คำทั้งหมดที่มีในแบบทดสอบการรู้คิดแบบ Recall ของโครงการ HART จำแนกตามชุดคำ

ชุดคำ	คำ									
	1	2	3	4	5	6	7	8	9	10
1	โรงแรม	แม่น้ำ	โต๊ะ	ผิวหนัง	ทอง	ตลาด	กระต่าย	เด็ก	ในหลวง	ไฟ
2	ท้องฟ้า	ยา	ธงชาติ	ข้าว	ภรรยา	เครื่องจักร	บ้าน	ดิน	มะพร้าว	ตำบล
3	สตรี	ก้อนหิน	เลือด	หมอ	รองเท้า	จดหมาย	พระ	สวน	หุบเขา	เครื่องยนต์
4	น้ำ	วัด	พยาบาล	ปราสาท	ฝน	ดอกไม้	ทะเล	หมู่บ้าน	ผู้ชาย	ญาติ

ตัวแปรหลักและค่าของตัวแปรหลักสำหรับการศึกษาคครั้งนี้เป็นดังนี้

1. จำนวนผู้ที่ตอบแบบทดสอบการรู้คิดแบบ Recall ด้วยชุดคำที่ k (N_k) โดย $N_1 = 1,600$ คน

$N_2 = 1,030$ คน $N_3 = 879$ คน $N_4 = 803$ คน และ $\sum N_k = 4,312$ คน

2. คะแนนที่จำคำที่ i ได้ของคนที่ j ในชุดคำที่ k (x_{ijk}) โดย $x_{ijk} = 0$ ถ้าจำไม่ได้ และ $x_{ijk} = 1$ ถ้าจำได้

ดังนั้น $\sum_i x_{ijk}$ = จำนวนผู้ที่ตอบแบบทดสอบการรู้คิดแบบ Recall ด้วยชุดคำที่ k และจำคำที่ i ได้

$\Sigma_j x_{kij}$ = จำนวนคำผู้ตอบแบบทดสอบการรู้คิดแบบ Recall ด้วยชุดคำที่ k คนที่ j จำได้

4.2 การวิเคราะห์ข้อมูล

การประเมินประสิทธิผลของแบบทดสอบการวัดการรู้คิดแบบ Recall ของโครงการ HART พิจารณาจากระดับความยาก อำนาจการจำแนก ความเชื่อถือได้ตามสูตรการคำนวณที่ระบุไว้ในหัวข้อ 3.2 และใช้สถิติเชิงพรรณนา (Descriptive Statistics) ในรูปแบบของร้อยละ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน รวมทั้งการนำเสนอตัวชี้วัดด้วยแผนภูมิเส้น (Line Chart) และแผนภาพการกระจาย (Scatter Diagram) ส่วนการเปรียบเทียบประสิทธิผลของแบบทดสอบการวัดการรู้คิดแบบ Recall ระหว่างชุดคำต่าง ๆ ทั้ง 4 ชุดว่ามีระดับความยาก อำนาจการจำแนก แตกต่างกันหรือไม่อย่างไรนั้นจะใช้การวิเคราะห์ความแปรปรวนทางเดียว (One-Way Analysis of Variance: One-Way ANOVA) โดยมีสมมติฐานทางการวิจัยดังนี้

สมมุติฐานที่ 1: ระดับความยากของชุดคำแตกต่างกันหรือไม่ อย่างไร

สมมุติฐานที่ 2: อำนาจการจำแนกของชุดคำแตกต่างกันหรือไม่ อย่างไร

5. ผลการศึกษา

5.1 การวิเคราะห์ข้อมูลเบื้องต้น

ในการสำรวจรอบ (Wave) ที่ 4 ของโครงการ HART พบว่ามีผู้ให้ความร่วมมือทำแบบทดสอบการรู้คิด (Cognition) ทั้งหมด 4,312 คน พบว่าเมื่อพิจารณาจำนวนคำที่คนจำได้น้อย (น้อยกว่าร้อยละ 20) พบว่ามีอยู่ในชุดคำที่ 1 2 และ 3 ชุดละ 1 คำ ส่วนชุดคำที่ 4 ไม่มีลักษณะนี้เลย ในทางกลับกันไม่มีชุดคำไหนเลยที่มีคำที่คนจำได้มากกว่าร้อยละ 70 ยกเว้นชุดคำที่ 1 ซึ่งมี 1 คำที่มีคนจำได้ถึงร้อยละ 73 ในภาพรวมพบว่าร้อยละของคนที่ยังจำคำต่าง ๆ ในชุดคำที่ 1 ได้มีค่าประมาณร้อยละ 18 – 73 ส่วนชุดคำที่ 2 – 4 นั้นคนที่จำคำต่าง ๆ ในแต่ละชุดคำมีค่าประมาณร้อยละ 19 – 59 10 – 64 และ 23 – 63 ตามลำดับ ดังรายละเอียดแสดงไว้ในตารางที่ 5

ตารางที่ 5 จำนวนและร้อยละ (ค่าในวงเล็บ) ของผู้ที่จำคำต่าง ๆ ของชุดคำทั้ง 4 ชุด ในแบบทดสอบการรู้คิดแบบ Recall ของโครงการ HART สำหรับการสำรวจรอบ (Wave) ที่ 4

ชุดคำ	คำ										รวม*
	1	2	3	4	5	6	7	8	9	10	
1	960 (60.00)	976 (61.00)	597 (37.31)	291 (18.19)	568 (35.50)	635 (39.69)	595 (37.19)	679 (42.44)	1,172 (73.25)	875 (54.69)	1,600
2	590 (57.28)	403 (39.13)	199 (19.32)	477 (46.31)	334 (32.43)	255 (24.76)	538 (52.23)	509 (49.42)	485 (47.09)	611 (59.32)	1,030
3	502 (57.11)	328 (37.32)	323 (36.75)	330 (37.54)	210 (23.89)	91 (10.35)	509 (57.91)	277 (31.51)	417 (47.44)	560 (63.71)	879
4	508 (63.26)	517 (64.38)	183 (22.79)	290 (36.11)	255 (31.76)	232 (28.89)	392 (48.82)	249 (31.01)	371 (46.20)	472 (58.78)	803

* จำนวนผู้ที่ตอบแบบทดสอบการรู้คิดแบบ Recall ด้วยชุดคำที่ k (k = 1, 2, 3, 4)

สำหรับผลการศึกษาจะแนรวมของผู้สูงอายุซึ่งแบ่งออกเป็น 4 ช่วงอายุ พบว่า โดยภาพรวมผู้สูงอายุทั้ง 4 กลุ่มได้คะแนนความจำในลักษณะเดียวกัน คือ ผู้สูงอายุได้คะแนนรวม 3 – 5 คะแนน คิดเป็นร้อยละ 66 ผู้สูงอายุได้คะแนนรวม 6 – 10 คะแนน คิดเป็นร้อยละ 22 และผู้สูงอายุได้คะแนนรวม 0 – 2 คะแนน คิดเป็นร้อยละ 12 ดังรายละเอียดแสดงไว้ในตารางที่ 6

ตารางที่ 6 จำนวนและร้อยละ (ค่าในวงเล็บ) ของคะแนนรวมของผู้ที่จำคำต่าง ๆ ของชุดคำ 4 ชุดในแบบทดสอบการรู้คิดแบบ Recall ของโครงการ HARTจากการสำรวจรอบ (Wave) ที่ 4

คะแนนรวม	ช่วงอายุ				
	45-59 ปี	60-69 ปี	70-79 ปี	80 ปีขึ้นไป	รวม
0	17	20	14	14	65 (1.51)
1	33	34	30	26	123 (2.85)
2	88	107	91	44	330 (7.65)
3	257	282	222	130	891 (20.66)
4	368	376	260	145	1,149 (26.65)
5	255	248	174	115	792 (18.37)
6	185	141	101	54	481 (11.15)
7	86	69	62	23	240 (5.57)
8	31	28	21	11	91 (2.11)
9	9	13	10	9	41 (0.95)
10	37	34	21	17	109 (2.53)
รวม	1,366	1,352	1,006	588	4,312

5.2 ประสิทธิภาพของแบบทดสอบการวัดการรู้คิดแบบ Recall ของโครงการ HART

ผลการศึกษา แบ่งออกเป็น 3 ส่วนหลักได้แก่ การศึกษาระดับความยาก อำนาจการจำแนก และความเชื่อถือได้ ซึ่งมีรายละเอียด ดังต่อไปนี้

5.2.1 ระดับความยาก (Difficulty Level)

ผลการศึกษาในระดับความยาก พบว่า ชุดคำทั้ง 4 เป็นชุดคำที่อยู่ในระดับยากปานกลาง นอกจากนี้ไม่มีคำที่ง่ายมากอยู่เลย หากพิจารณาในรายละเอียด จะเห็นว่า ทุกชุดคำมีคำระดับยากปานกลางคิดเป็นร้อยละ 80 และชุดคำที่ 1 2 และ 3 มีคำที่ควรนำไปพิจารณาปรับปรุง 1 คำ และไม่ควรนำไปใช้ 1 คำ ส่วนชุดคำที่ 4 มีคำที่ควรนำไปพิจารณาปรับปรุง 2 คำ (ตารางที่ 7 และ ตารางที่ 8)

ตารางที่ 7 ระดับความยาก (Difficulty Level) ของแต่ละคำจากชุดคำทั้ง 4 ชุด

ระดับความยาก	ความหมาย	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4
0.00 – 0.20	ยากมาก	ผิวหนัง**	ธงชาติ**	จดหมาย**	-
0.21 – 0.30	ยาก	-	เครื่องจักร*	รองเท้า*	พยาบาล* ดอกไม้
0.31 – 0.70	ปานกลาง	โรงแรม แม่น้ำ โต๊ะ ทอง ตลาด กระดาษ เด็ก ไฟ	ท้องฟ้า ยา ข้าว กรรยบ้าน ดิน มะพร้าว ตำบล	สตรี ก้อนหิน เลือด หมอ พระ สวน หุบ เขา เครื่องยนต์	น้ำ วัด ปราสาท ฝน ทะเล หมู่บ้าน ผู้ชาย ญาติ
0.71 – 0.89	ง่าย	ในหลวง*	-	-	-
0.90 – 1.00	ง่ายมาก	-	-	-	-

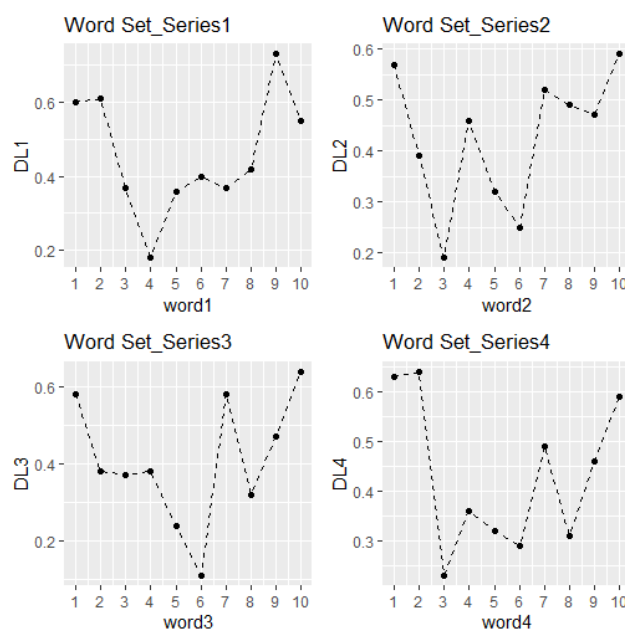
*ควรพิจารณาปรับปรุง

** ไม่นำไปใช้

ตารางที่ 8 ค่าประมาณแบบช่วงของค่าเฉลี่ยระดับความยาก (Difficulty Level: DL) ของชุดคำทั้ง 4 ชุด

ชุดที่	ค่าเฉลี่ย (ส่วนเบี่ยงเบน มาตรฐาน)	ความหมาย	95% C.I. of mean of DL	คำที่มีระดับ ความยากสูงสุด (DL)	คำที่มีระดับ ความยากต่ำสุด (DL)
1	0.46 (0.16)	ยากปานกลาง	(0.34 - 0.57)	ผิวหนัง (0.18)	ในหลวง (0.73)
2	0.43 (0.14)	ยากปานกลาง	(0.33 - 0.53)	ธงชาติ (0.19)	ตำบล (0.59)
3	0.40 (0.17)	ยากปานกลาง	(0.29 - 0.52)	จดหมาย (0.11)	เครื่องยนต์ (0.64)
4	0.43 (0.15)	ยากปานกลาง	(0.32 - 0.54)	ดอกไม้ (0.29)	วัด (0.64)

เมื่อพิจารณาแผนภูมิเส้นระดับความยากในชุดคำแต่ละชุด (ภาพที่ 1) พบว่าภาพรวมของระดับความยากของชุดคำทั้ง 4 ชุดมีค่าเฉลี่ยและมีการกระจายของค่าระดับความยากใกล้เคียงกัน



ภาพที่ 1 แผนภูมิเส้นแสดงระดับความยาก(Difficulty Level)

หลังจากที่ได้ตรวจสอบข้อตกลงเบื้องต้นสำหรับการแจกแจงปกติด้วย Shapiro-Wilk Test และความเป็นเอกพันธ์ของความแปรปรวน (Homogeneity of Variance) ของด้วย Levene 's Test ผ่านเรียบร้อยแล้ว จึงได้ทำการทดสอบความแตกต่างค่าเฉลี่ยระดับความยากของชุดคำทั้ง 4 ชุด โดยใช้การวิเคราะห์ความแปรปรวนทางเดียว (One way Analysis of Variance : ANOVA) ที่ระดับนัยสำคัญ 0.05 ได้ผลการวิเคราะห์เป็นดังนี้

ตารางที่ 9 ANOVA Table สำหรับสมมติฐานที่ 1

	Sum of Squares	df	Mean Square	F	P-value
Between Groups	0.0157	3	0.0052	0.221	0.881
Within Groups	0.8522	36	0.0237		
Total	0.8679	39			

ในการทดสอบสมมติฐานทางที่ 1 ($P\text{-value} = 0.881 > \alpha = 0.05$) พบว่าความแตกต่างของค่าเฉลี่ยระดับความยากของชุดคำทั้ง 4 ชุดไม่มีนัยสำคัญทางสถิติ

5.2.2 อำนาจการจำแนก (Discriminating Power)

ผลการศึกษาพบว่า ชุดคำทั้ง 4 เป็นชุดคำที่มีอำนาจการจำแนกดีมาก หากพิจารณาในรายละเอียดจะเห็นว่า ชุดคำที่ 1 เป็นเพียงชุดคำเดียวที่มีค่าอำนาจการจำแนกดีมาก คิดเป็นร้อยละ 80 ในขณะที่ชุดคำที่ 2 3 และ 4 มีค่าอำนาจการจำแนกดีมาก คิดเป็นร้อยละ 60 โดยชุดคำที่ 1 มีคำที่ควรนำไปปรับปรุงเมื่อมีโอกาส 1 คำ และคำที่ต้องทบทวนการใช้ 1 คำ ส่วนชุดคำที่ 2 และ 3 มีคำที่ควรนำไปปรับปรุงเมื่อมีโอกาส 2 คำ และคำที่ต้องทบทวนการใช้ 2 คำ และชุดคำที่ 4 มีคำที่ควรนำไปปรับปรุงเมื่อมีโอกาส 3 คำ และคำที่ต้องทบทวนการใช้ 1 คำ (ตารางที่ 10 ตารางที่ 11 และ ตารางที่ 12)

ตารางที่ 10 จำนวนผู้สูงอายุที่ตอบถูกต้องในกลุ่มบน (R_U) และกลุ่มล่าง(R_L)

ชุด		คำ										ร้อยละ 27 ของ ผู้สูงอายุทั้งหมด
		1	2	3	4	5	6	7	8	9	10	
1	R_U	386	365	326	198	257	278	315	292	386	292	432
	R_L	138	142	43	15	55	75	49	71	237	175	432
2	R_U	231	188	103	177	178	136	198	194	163	192	278
	R_{Lr}	90	47	19	61	26	24	85	92	108	126	278
3	R_U	204	157	169	142	119	57	171	123	165	171	237
	R_L	82	58	18	40	13	7	88	36	64	115	237
4	R_U	192	185	98	140	113	121	159	109	136	154	217
	R_L	88	82	17	25	26	20	55	39	59	111	217

ตารางที่ 11 อำนาจการจำแนก (Discriminating Power: DP) ของแต่ละคำจากชุดคำทั้ง 4 ชุด

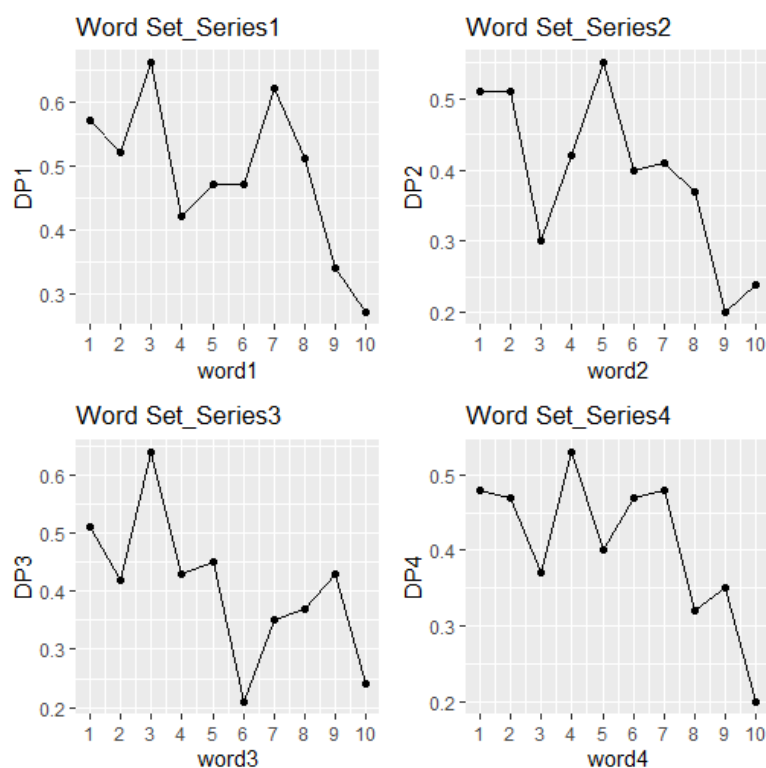
DP	ความหมาย	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4
> 0.39	ดีมาก	โต๊ะ กระดาษ โรงแรม แม่น้ำ เด็ก ทอง ตลาด ผิวหนัง	ภรรยา ท้องฟ้า ยา ข้าว เครื่องจักร บ้าน	เลือด สตรี รองเท้า หมอ ก้อนหิน หุบเขา	ปราสาท น้ำ วัด ทะเล ดอกไม้ ฝน
0.30-0.39	ดี	ในหลวง*	ดิน* ธงชาติ*	สวน* พระ*	พยาบาล* ผู้ชาย* หมูบ้าน*
0.20-0.29	ปานกลาง	ไฟ**	ตำบล* มะพร้าว**	เครื่องยนต์** จดหมาย**	ญาติ**
0.00-0.19	ค่อนข้างไม่ดี	-	-	-	-
< -0.01	ไม่ดี	-	-	-	-

** ทบทวนการใช้ *ควรนำไปปรับปรุงเมื่อมีโอกาส

ตารางที่ 12 ค่าประมาณแบบช่วงของค่าเฉลี่ยอำนาจการจำแนก (Discriminating Power: DP) ของชุดคำทั้ง 4 ชุด

ชุดที่	ค่าเฉลี่ย (ส่วนเบี่ยงเบน มาตรฐาน)	ความหมาย	95% C.I. of mean of DP	ค่าที่มีอำนาจการ จำแนกสูงสุด (DP)	ค่าที่มีอำนาจการ จำแนกต่ำสุด (DP)
1	0.49 (0.12)	ดีมาก	(0.40 - 0.57)	โต๊ะ (0.66)	ไฟ (0.27)
2	0.39 (0.12)	ดีมาก	(0.31 - 0.48)	ภรรยา (0.55)	มะพร้าว (0.20)
3	0.41 (0.12)	ดีมาก	(0.32 - 0.49)	เลือด (0.64)	จดหมาย (0.21)
4	0.41 (0.10)	ดีมาก	(0.34 - 0.49)	ปราสาท (0.53)	ญาติ (0.20)

เมื่อพิจารณาแผนภูมิเส้นอำนาจการจำแนกในชุดคำแต่ละชุด (ภาพที่ 2) พบว่าชุดคำที่ 1 ก่อนข้างมีอำนาจการการจำแนกสูงกว่าชุดอื่นๆ ในขณะที่ชุดคำที่ 2 3 และ 4 มีอำนาจการจำแนกใกล้เคียงกัน นอกจากนี้ยังพบว่าชุดคำทั้ง 4 ชุดมีการกระจายของค่าอำนาจการจำแนกใกล้เคียงกัน



ภาพที่ 2 แผนภูมิเส้นอำนาจจำแนกของคำ (Discrimination Power)

หลังจากที่ได้ตรวจสอบข้อตกลงเบื้องต้นสำหรับการแจกแจงปกติด้วย Shapiro-Wilk Test และความเป็นเอกพันธ์ของความแปรปรวน (Homogeneity of Variance) ของด้วย Levene 's Test ผ่านเรียบร้อยแล้ว จึงได้ทำการทดสอบความแตกต่างค่าเฉลี่ยอำนาจการจำแนกของชุดคำทั้ง 4 ชุด โดยใช้การวิเคราะห์ความแปรปรวนทางเดียว (One way Analysis of Variance : ANOVA) ที่ระดับนัยสำคัญ 0.05 ได้ผลการวิเคราะห์เป็นดังนี้

ตารางที่ 13 ผลการทดสอบสมมติฐานที่ 2

	Sum of Squares	df	Mean Square	F	P-value
Between Groups	0.0524	3	0.01807	1.354	0.272
Within Groups	0.4802	36	0.01334		
Total	0.5344	39			

ในการทดสอบสมมติฐานทางที่ 2 ($P\text{-value} = 0.272 > \alpha = 0.05$) พบว่าความแตกต่างของค่าเฉลี่ยอำนาจการจำแนกของชุดคำทั้ง 4 ชุดไม่มีนัยสำคัญทางสถิติ

5.2.3 ความสัมพันธ์ระหว่างระดับความยากและอำนาจการจำแนก

เมื่อพิจารณาแผนภาพการกระจายแสดงความสัมพันธ์ระหว่างระดับความยากและอำนาจการจำแนก (ภาพที่ 3) พบว่า ระดับความยากและอำนาจการจำแนกไม่มีความสัมพันธ์เชิงเส้นในทุกชุดคำ



ภาพที่ 3 แผนภาพการกระจายของระดับความยากและอำนาจการจำแนกของชุดคำทั้ง 4 ชุด

5.2.4 ความเชื่อถือได้ (Reliability)

สำหรับค่าความเชื่อถือได้ (Reliability) ของ Kuder-Richardson Formula 20 (KR-20) พบว่า ชุดคำที่ 1 มีค่าความเชื่อถือไม่ได้ ในขณะที่ชุดคำที่ 2 3 และ 4 มีค่าความเชื่อถือได้อยู่ในเกณฑ์ที่ไม่สามารถยอมรับได้

ตารางที่ 14 ความน่าเชื่อถือได้ (Reliability) ของชุดคำทั้ง 4 ชุด

ชุดคำ	1	2	3	4
ความเชื่อถือได้ (Reliability)	0.51	0.25	0.35	0.31

6. สรุปและอภิปรายผลการศึกษา

การศึกษานี้มีวัตถุประสงค์เพื่อประเมินและเปรียบเทียบประสิทธิผลของแบบทดสอบการวัดการรู้คิดแบบ Recall ที่ใช้กับผู้สูงอายุจากโครงการสำรวจด้านสุขภาพ การสูงอายุ และการเกษียณในประเทศไทย (HART) โดยใช้ข้อมูลจากการสำรวจรอบที่ 4 ที่ดำเนินการในปี พ.ศ. 2565-2566 โดยพิจารณาจากระดับความยาก (Difficulty Index) อำนาจการจำแนก (Discriminating Index) ความเชื่อถือได้ (Reliability) ของแบบทดสอบทั้ง 4 ชุดซึ่งแต่ละชุดประกอบด้วยคำ 10 คำ ทางโครงการฯ ได้จัดทำแบบทดสอบการจำคำได้สำหรับผู้สูงอายุไว้จำนวน 4 ชุด แบบทดสอบ 1 ชุดซึ่งมีคำอยู่จำนวน 10 คำ ได้ถูกสุ่มให้แก่ผู้สูงอายุแต่ละคน ผลการศึกษาพบว่า จำนวนคำที่คนจำได้น้อย (น้อยกว่าร้อยละ 20) มีอยู่ในชุดคำที่ 1 2 และ 3 ชุดละ 1 คำ ส่วนชุดคำที่ 4 ไม่มีคำลักษณะนี้เลย ในทางกลับกันไม่มีชุดคำไหนเลยที่มีคำที่คนจำได้มากกว่าร้อยละ 70 ยกเว้นชุดคำที่ 1 ซึ่ง

มี 1 คำที่มีคนจำได้ถึงร้อยละ 73 โดยภาพรวมผู้สูงอายุทั้ง 4 ช่วงอายุได้คะแนนความจำในลักษณะเดียวกัน คือ ได้คะแนนรวม 3 - 5 คะแนน คิดเป็นร้อยละ 66 โดยผู้สูงอายุส่วนใหญ่ได้คะแนนรวม 4 คะแนน

ในภาพรวมชุดคำทั้ง 4 มีคําระดับยากปานกลางคิดเป็นร้อยละ 80 ชุดคำที่ 1 มีคำที่มีอำนาจการจำแนกดีมาก คิดเป็นร้อยละ 80 ในขณะที่ชุดคำที่ 2 3 และ 4 มีคำที่มีอำนาจจำแนกดีมากเท่ากันคือ ร้อยละ 60 (ตารางที่ 15) นอกจากนี้ยังพบว่าความแตกต่างของค่าเฉลี่ยระดับความยากและอำนาจการจำแนกของชุดคำทั้ง 4 ชุดไม่มีนัยสำคัญทางสถิติ และระดับความยากและอำนาจการจำแนกไม่มีความสัมพันธ์เชิงเส้นต่อกัน ในส่วนของความเชื่อถือได้ พบว่าชุดคำที่ 1 มีความเชื่อถือได้ไม่ดี ในขณะที่ชุดคำอีก 3 ชุด มีความเชื่อถือได้อยู่ในเกณฑ์ที่ไม่สามารถยอมรับได้

ตารางที่ 15 สรุปผลการศึกษาระดับความยาก (Difficulty Level: DL) และอำนาจการจำแนก (Discriminating Power : DP) ของชุดคำทั้ง 4 ชุด

ชุดที่	ระดับความยาก	คำแนะนำ	อำนาจการจำแนก	คำแนะนำ
1	ปานกลาง	ผิวหนัง (ไม่ใช่)	ดีมาก	ในหลวง (ควรปรับปรุงเมื่อมีโอกาส) ไฟ (ทบทวนการใช้)
2	ปานกลาง	ธงชาติ (ไม่ใช่) เครื่องจักร (ปรับปรุง)	ดีมาก	ดิน ธงชาติ (ควรปรับปรุงเมื่อมีโอกาส) มะพร้าว ตำบล (ทบทวนการใช้)
3	ปานกลาง	จดหมาย (ไม่ใช่) รองเท้า (ปรับปรุง)	ดีมาก	สวน พระ (ควรปรับปรุงเมื่อมีโอกาส) เครื่องยนต์ จดหมาย (ทบทวนการใช้)
4	ปานกลาง	พยาบาล (ปรับปรุง) ดอกไม้ (ปรับปรุง)	ดีมาก	พยาบาล ผู้ชาย หมู่บ้าน (ควรปรับปรุงเมื่อมีโอกาส) ญาติ (ทบทวนการใช้)

ผลการวิเคราะห์จากงานวิจัยนี้ประกอบกับผลการศึกษาเรื่อง Word-list Recall ในคนอายุน้อย (น้อยกว่า 20 ปี) และคนอายุมาก (มากกว่า 50 ปี) ของ Khosravizadeh & Gerami, 2011 ซึ่งกล่าวว่า ความสามารถในการจดจำระยะสั้น (Short-term Memory) ของคนจะลดลงเมื่อมีอายุเพิ่มขึ้น และ หากความหมายของคำศัพท์นั้นมีความเกี่ยวข้องหรือมีความสัมพันธ์กันกับคำศัพท์อื่นในชุดคำเดียวกันจะทำให้คนจดจำคำนั้นได้มากขึ้น ทำได้ข้อสรุปเพิ่มเติมที่เป็นประโยชน์ 2 ประเด็น ประเด็นแรกคือ สำหรับผู้สูงอายุทั้ง 4 ช่วงอายุความสามารถในการจดจำคำมีลักษณะแบบเดียวกันทุกกลุ่ม นั่นคือผู้สูงอายุทั้ง 4 ช่วงอายุได้คะแนนรวม 3 - 5 คะแนน คิดเป็นร้อยละ 66 โดยผู้สูงอายุส่วนใหญ่ได้คะแนนรวม 4 คะแนน และประเด็นที่ 2 คือ ความเชื่อถือได้ซึ่งพิจารณาจากความสม่ำเสมอภายใน (Internal Consistency) ของคำศัพท์แต่ละคำในแต่ละชุดคำ พบว่า ชุดคำที่ 1 มีความเชื่อถือได้ไม่ดี ในขณะที่ชุดคำอีก 3 ชุด มีความเชื่อถือได้อยู่ในเกณฑ์ที่ไม่สามารถยอมรับได้ ซึ่งหากต้องการปรับปรุงความเชื่อถือได้ของชุดคำทั้ง 4 ชุดควรนำประเด็นเรื่องคำศัพท์ที่มีความหมายเกี่ยวข้องกันกับคำศัพท์อื่นๆในชุดคำเดียวกันมาพิจารณาร่วมด้วย เพื่อหลีกเลี่ยงความเอนเอียง (Bias) ที่อาจส่งผลให้คนจำคำนั้นได้มากขึ้น

7. ข้อเสนอแนะ

ในการประเมินและเปรียบเทียบประสิทธิผลของแบบทดสอบการวัดการรู้คิดแบบ Word Recall ได้ศึกษากับผู้สูงอายุที่ถูกแบ่งออกเป็น 4 ช่วงอายุ ซึ่งในความเป็นจริงแล้วยังมีตัวแปรสำคัญอื่นๆ เช่น ระดับการศึกษาสูงสุด พฤติกรรมการรับประทานอาหาร การนอนหลับ การออกกำลังกาย การดื่มแอลกอฮอล์ ของผู้สูงอายุที่ควรนำมาศึกษา ร่วมด้วยเพื่อเพิ่มประสิทธิผลจากการใช้แบบทดสอบการรู้คิดแบบ Word Recall นอกจากนี้ หน่วยงานทั้งของภาครัฐ ได้แก่ กรมกิจการผู้สูงอายุ ภายใต้กำกับกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ กระทรวงสาธารณสุข กระทรวงแรงงาน ฯลฯ ตลอดจนภาคเอกชน สามารถนำผลวิจัยนี้ไปใช้ให้เกิดประโยชน์ ในการส่งเสริมและพัฒนาศักยภาพของผู้สูงอายุในประเทศไทย

8. เอกสารอ้างอิง

- กรมกิจการผู้สูงอายุ. สถิติผู้สูงอายุ กันยายน 2567. ค้นเมื่อ 5 กุมภาพันธ์ 2568 จาก <https://www.dop.go.th/th/statistics>.
- สำนักข่าว Hfocus. กรมการแพทย์ห่วง “สังคมสูงวัย” พบ! ผู้สูงอายุป่วยสมองเสื่อมเพิ่มขึ้นทุกปี. ค้นเมื่อ 5 กุมภาพันธ์ 2568 จาก <https://www.hfocus.org/content/2023/12/29302>.
- Alzheimer's Association. (2017). 2017 Alzheimer's Disease Facts and Figures. **Alzheimer's & Dementia**, 13(4), 325-373. doi:10.1016/j.jalz.2017.02.001.
- Backhoff, E., Larrazolo, N., & Rosas, M. (2000). The Level of Difficulty and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA). **Revista Electrónica de Investigación Educativa**, 2(1), 1 – 16.
- Basuki, I. & Hariyanto, D. (2015). **Assesmen Pembelajaran**. Bandung: Remaja Rosdakarya.
- Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H.T.V., & Croteau, J. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer 's type dementia in older adults: a systematic review and meta-analysis. **Neuropsychology Review**, 27, 328-353. doi:10.1007/s11065-017-9361-5.
- Brown, A. (2019). The Role of Cognitive Testing in the Elderly: Insights and Implications. **Aging and Mental Health**, 23(8), 1021-1030. doi:10.1080/13607863.2019.1574298.
- CERAD (2025). **Clinical Assessment Battery for Alzheimer 's Disease (electronic form)**. Retrieved 5 February 2025 from <https://agingcenter.duke.edu/cerad>.
- Considine, J., Botti, M., & Thomas, S. (2005). Design, Format, Validity and Reliability of Multiple Choice Questions for Use in Nursing Research and Education. **Collegian**, 12(1), 19-24. doi:10.1016/S1322-7696(08)60478-3.
- Episodic Memory (2024). Retrieved 5 February 2025 from https://en.wikipedia.org/wiki/Episodic_memory.
- Freeman, F. S. (1962). **Theory and Practice of Psychological Testing**. (3rd ed.). New York, NY: Holt, Rinehart and Winston. ISBN 978-0039102855.
- Friatma, A., & Anhar, A. (2019). Analysis of validity, reliability, discrimination, difficulty and distraction in effectiveness assessment. **Journal of Physics: Conference Series**, 1387, 1-6. doi:10.1088/1742-6596/1387/1/012063.

- Gronlund, N. E. (1993). **Measurement and Evaluation in Teaching**. (6th ed.). New York, NY: Macmillan Publishing Company. ISBN 978-0023481112.
- Hunziker, M., & Hass, L. (2014). **Building Blocks of Personality Type: A Guide to Exploring the Eight Jungian Functions**. Telos Publications.
- Jung, C. G. (1971). **Psychological Types (H. G. Baynes, Trans.)**. Princeton University Press.
- Khosravizadeh, P., & Gerami, P. (2006). Word-list Recall in Youngsters and Older Adults. **Brain. Board Research in Artificial Intelligence and Neuroscience**, 2(1), 5-10. ISSN 2067-3957.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. **Psychometrika**, 2(3), 151-160.
- Lange, P. V., & Mehrens, W. A. (1967). **Measuring Educational Achievement**. Englewood Cliffs, NJ: Prentice-Hall. ISBN 978-0135669999.
- Myers, I.B., & Myers, P.B. (2010). **Gifts Differing: Understanding Personality Type**. Mountain View, CA: Nicholas Brealey.
- Popham, W. J. (2002). ***Classroom Assessment: What Teachers Need to Know***. (3rd ed.). Boston, MA: Allyn & Bacon. ISBN 978-0205349803.
- Sharma, S. (2000). **Educational Measurement and Evaluation**. New Delhi: Commonwealth Publishers. ISBN 978-8171695692.
- Singh, A., Gupta, R., & Singh, S. (2009). **Educational Measurement and Evaluation**. New Delhi: APH Publishing Corporation. ISBN 978-8131304887.
- Sim, S. M., & Rasiyah, R. I. (2006). Relationship between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. **Annals of the Academy of Medicine**, 35(2), 67-71. ISSN 0304-4602.
- Smith, J., Brown, T., Johnson, A., & Roberts, M. (2018). Cognitive Assessment and Memory Decline in the Elderly. **Journal of Aging and Health**, 30(5), 679-694. doi:10.1177/0898264317698552.
- Trice, A. D. (2000). **A Handbook of Classroom Assessment**. Long Grove, IL: Waveland Press. ISBN 978-1577661675.
- Wagle, J., Selbaek, G., Benth, J.S., Gjora, L., Ronqvist, T.K., Bekkhus-Wetterberg, P., Persson, K., & Engedal, K. (2023). The CERAD Word List Memory Test: Normative Data Based on a Norwegian Population-Based Sample of Healthy Older Adults 70 Years and Above. The HUNT Study. **Journal of Alzheimer Disease**, 91(1), 321-343. doi:10.3233/JAD-220672.
- World Health Organization. (2023). **Dementia**. Retrieved 5 February 2025 from <https://www.who.int/news-room/fact-sheets/detail/dementia>.
- Zubairi, A. M., & Kassim, N. L. A. (2006). **Classroom Test and Item Analysis: An Overview of Measurement Issues**. In Proceedings of the 2nd International Conference on Measurement and Evaluation in Education (ICMEE 2006), Malaysia.