



บทที่ 1

บทนำ

### ความเป็นมาและความสำคัญของปัญหา

แบบล่อทางจิตวิทยาและการศึกษาเป็นเครื่องมือที่เราใช้เพื่อให้ได้กลุ่มตัวอย่าง พฤติกรรมของบุคคลซึ่งมีการทำให้เป็นปริมาณเพื่อให้ได้คะแนนเป็นตัว เลขที่นำไปเป็นแนว ทางสรุปคุณลักษณะบางประการภายในบุคคลนั้น (Latent trait) และเมื่อแบบล่อมี ลักษณะเป็นแบบเลือกตอบแล้ว วิธีตรวจให้คะแนนที่นิยมปฏิบัติกันโดยทั่วไป ตามทฤษฎี มาตรฐานเดิม (Classical Test Theory) คือ การให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) โดยให้คะแนน 1 สำหรับการตอบถูก และให้ 0 สำหรับการ ตอบผิด หรือไม่ตอบในข้อนั้น ๆ แล้วตัดสิ้นผลการล่อจากผลรวมของคะแนนรายข้อนั้น เนื่องจากมั่นใจว่าเป็นวิธีที่ดีและสะดวกทั้งคะแนนที่ได้รับก็สามารถเป็นตัว แทนที่ดีตั้งค่ากล่าว ของ กุลลิคเสน (Gulliksen 1967 : 247) ว่า "ในกรณีที่สามารถทำให้การล่อมี ลักษณะเป็น Power test แล้ว คะแนนที่ได้จากจำนวนข้อกระทงที่ตอบถูกเป็นคะแนนที่ สามารถนำไปใช้ได้ดี สะดวก รวดเร็ว โดยมีความคลาดเคลื่อนอย่างแท้จริงเพียง เล็กน้อย" แต่การให้คะแนนโดยวิธีนี้ก็ทำให้มีข้อสงสัยว่าอาจไม่ยุติธรรมสำหรับผู้ล่อ เนื่องจากคะแนนเหล่านั้นขึ้นอยู่กับลักษณะของข้อกระทงเป็นสำคัญ (Angoff 1971 อ้างถึง ใน ชูศักดิ์ ชัมภลสิทธิ์ 2525 : 34) โดยเฉพาะสำหรับข้อกระทงที่ไม่เคยทดลองใช้มาก่อน และเนื่องจากปัจจุบันพบว่าทฤษฎีมาตรฐานเดิม (Classical Test Theory) เป็นโมเดล ที่อ่อน (weak model) และมีจุดอ่อนหลายประการ (Hambleton and Liden 1982 : 373 - 375) คือ ประการแรก ค่าพารามิเตอร์ของข้อกระทง คือ ค่าความยาก และ ค่าอำนาจจำแนก แปรเปลี่ยนไปตามกลุ่มผู้ล่อที่แตกต่างกันในด้านความสามารถ (ability) หมายความว่า ค่าสถิติของข้อกระทงดังกล่าวมีประโยชน์ในการสร้างแบบล่อก็เฉพาะแต่ กลุ่มประชากรผู้ล่อที่มีลักษณะคล้ายคลึงกับกลุ่มตัวอย่างผู้ล่อที่ใช้ในการหาค่าสถิติเหล่านั้น

ซึ่งในหลาย ๆ สถานการณ์ของการสอบ เราต้องการพารามิเตอร์ของข้อกระทงที่ไม่แปรเปลี่ยน จุดอ่อนประการที่สอง คือ การเปรียบเทียบผู้สอบในเรื่องของความสามารถที่ระดับหนึ่งซึ่งวัดจากแบบสอบ จำกัดอยู่กับสถานการณ์ที่ผู้สอบถูกทดสอบด้วย นั่นคือ ไขข้อกระทงชุดเดียวกันหรือที่เป็นคู่ขนานกัน ซึ่งส่วนใหญ่ของแบบสอบมาตรฐาน วัดผลสัมฤทธิ์และความถนัดเฉพาะกับนักเรียนที่มีความสามารถปานกลาง ซึ่งข้อกระทงเหล่านี้ไม่ให้การประมาณค่าความสามารถที่แน่นอนสำหรับนักเรียนที่มีความสามารถระดับสูงหรือต่ำ จุดอ่อนประการที่สาม คือ เทคนิคการสอบที่ใช้กันอยู่กว้างขวางทั่วไปไม่ได้ให้รากฐานเพื่อกำหนดว่าผู้สอบคนหนึ่งจะทำข้อกระทงได้เพียงใด เมื่อได้เผชิญกับข้อกระทงหนึ่ง ยกเว้นเมื่อได้มีการใช้ข้อกระทงนั้นแล้วกับกลุ่มผู้สอบที่คล้ายคลึงกับบุคคลนั้น ในทางปฏิบัติเราต้องการพารามิเตอร์ของข้อกระทง (item parameters) และพารามิเตอร์ของผู้สอบ (examinee parameters) มาใช้เป็นตัวกำหนดลักษณะของข้อกระทง เพื่อที่จะสามารถทำนายความน่าจะเป็นที่ผู้สอบคนหนึ่งจะตอบข้อกระทงหนึ่งได้ถูก แม้แต่เมื่อข้อกระทงที่คล้ายคลึงกันนั้นยังไม่เคยใช้กับกลุ่มผู้สอบที่คล้ายคลึงกับผู้ตอบนี้เลย

จากจุดอ่อนทั้งสามประการและปัญหาอื่น ๆ ที่ทฤษฎีมาตรฐานเดิม (Classical Test Theory) ไม่สามารถให้คำตอบที่พอใจ นักทดสอบทางจิตวิทยาและการศึกษาคำนวณมากขึ้นได้สำรวจและพัฒนาทฤษฎีที่เหมาะสมกว่า ในการวัดความสามารถทางสติปัญญาและความพอใจส่วนใหญ่มุ่งไปที่ทฤษฎีการสนองตอบข้อกระทง (Item Response Theory: IRT หรือ Latent Trait Theory หรือ Item Characteristic Curve Theory หรือ Modern Test Theory) ซึ่งเริ่มด้วยโมเดลทางคณิตศาสตร์ที่กล่าวถึงความสัมพันธ์ระหว่างการตอบข้อกระทงของผู้สอบและระดับความสามารถของผู้สอบ ความสัมพันธ์นี้กำหนดได้ด้วยฟังก์ชันทางคณิตศาสตร์ที่มีชื่อเรียกว่า item characteristic function หรือเรียกว่า item characteristic curve (i.c.c.) สำหรับโมเดลทางคณิตศาสตร์ที่นำมาใช้อธิบายทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) นั้นมีหลายโมเดล แต่ละโมเดลแตกต่างกันด้วยรูปแบบของฟังก์ชันทางคณิตศาสตร์ที่ใช้และด้วยจำนวนของพารามิเตอร์ที่ใช้อธิบาย item characteristic function แต่ได้รับการพัฒนามากำหนดมาก และมีการนำไปใช้กันมากที่สุด ในหลาย ๆ สถานการณ์ของการสอบเนื่องจากสะดวกในการใช้

คือ Logistic Model ซึ่งแบ่งย่อยเป็น 3 โมเดล ตามจำนวนพารามิเตอร์ของฟังก์ชันดังกล่าว รูปสมการจะแสดงถึงความน่าจะเป็นของการตอบถูกหรือโอกาสตอบถูกที่ระดับความสามารถหนึ่ง ๆ จะตอบข้อกระทงแต่ละข้อได้ถูกต้อง ถ้าเป็นโมเดลที่มีพารามิเตอร์ 3 ตัว (Three - parameter Logistic Model : 3PL) รูปสมการก็จะใช้พารามิเตอร์ทั้ง 3 ตัว คือ  $a$  (ค่าอำนาจจำแนก),  $b$  (ค่าความยาก) และ  $c$  (ค่าการเดา) , ถ้าเป็นโมเดลที่มีพารามิเตอร์ 2 ตัว (Two-parameter Logistic Model : 2PL) มีข้อตกลงว่าไม่มีการเดา ( $c = 0$ ) , ดังนั้นจึงใช้พารามิเตอร์เพียง 2 ตัว คือ  $a$  และ  $b$  และถ้าเป็นโมเดลที่มีพารามิเตอร์เพียงตัวเดียว (One - parameter Logistic Model : 1PL หรือ Rasch Model) มีข้อตกลงว่าไม่มีการเดา ( $c = 0$ ) และค่าอำนาจจำแนกของทุกข้อกระทงเท่ากัน ( $\bar{a}$ ) พารามิเตอร์ที่ใช้จึงมี  $b$  เพียงตัวเดียว ทั้งสามโมเดลนี้กล่าวได้ว่า โมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) เป็นโมเดลที่มีความซับซ้อนมากที่สุด แต่ก็ยังเป็นโมเดลที่สอดคล้องกับลักษณะ และสถานการณ์ของการสอบที่เป็นจริงมากที่สุด (Thomas Warm 1978) ซึ่งจากผลการศึกษาของ Hambleton และ Traub (1973 อ้างถึงใน ผดุงจิต อินทสุวรรณ 2525 : 63) สรุปว่าโมเดลที่มีจำนวนพารามิเตอร์มากกว่าให้ความเหมาะสมกับข้อมูลดีกว่า

และภายใต้ทฤษฎีการสนองตอบข้อกระทง (IRT) นี้ได้เสนอวิธีตรวจให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามระดับความน่าจะเป็นของผู้สอบแต่ละคน ซึ่งการตรวจให้คะแนนแบบนี้จะทำให้ information ของข้อกระทงและของแบบสอบสูงสูดอันเป็นผลแสดงถึงความเชื่อถือได้ของแบบสอบนั้น ๆ หรือข้อกระทงนั้น ณ ระดับความสามารถต่าง ๆ กันของผู้สอบ สำหรับ Logistic Model นั้น น้ำหนักคะแนน (optimal weight) ขึ้นอยู่กับระดับความสามารถ ( $\theta$ ) นั่นคือระดับความสามารถสูง ๆ (การเดามีน้อยหรือไม่มีเลย) น้ำหนักคะแนน ( $w_i$ ) จะเป็นอิสระจากระดับความสามารถ แต่จะเป็นสัดส่วนโดยตรงกับค่าอำนาจจำแนก ( $a_i$ ) ส่วนที่ระดับความสามารถต่ำ ๆ นั้น น้ำหนักของข้อกระทงยาก ๆ เริ่มลดลง (มีการเดาเกิดขึ้น) และยิ่งความสามารถลดลงอีก น้ำหนักคะแนนของข้อกระทงจะเข้าใกล้ศูนย์ ด้วยวิธีดังกล่าวการให้คะแนนแบบนี้จึงน่าจะเป็นวิธีหนึ่งที่แก้ข้อสงสัยของการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) ได้ ผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบคะแนนสอบและคุณภาพของแบบ

ส่วจากการใช้วิธีตรวจให้คะแนนที่ต่างกันคือ การให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับ การให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) โดยเลือกโมเดลที่มีพารามิเตอร์ 3 ตัว (Three - parameter Logistic Model : 3PL) ศึกษาเปรียบเทียบกับคะแนนที่ได้รับจากแบบสอบ 3 วิชา คือ คณิตศาสตร์ วิทยาศาสตร์ และพละนาฏย ของนักเรียนจำนวนวิชาละ 3,907 คน ซึ่งได้จากการสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 141,306 คน ที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ของสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ โดยสุ่มแบบแบ่งชั้นภูมิ (Stratified Random Sampling) จากทุกเขตการศึกษา เขตการศึกษาละ 2.7% ของประชากรด้วยคอมพิวเตอร์ (SPSS)

เนื่องจากในประเทศไทยยังไม่มีผู้ใดศึกษาค้นคว้าในเรื่องนี้อย่างจริงจัง ผู้วิจัยคาดว่าผลการวิจัยครั้งนี้จะมีประโยชน์ในการเผยแพร่ทางวิชาการวัดผลการศึกษา ทำให้ทราบผลของการใช้วิธีการตรวจให้คะแนนที่แตกต่างกันที่มีต่อคะแนนสอบ เป็นแนวทางในการประยุกต์ใช้การตรวจให้คะแนนตามทฤษฎีการสนองตอบข้อกระทง (IRT) และเป็นแนวทางสำหรับผู้สนใจศึกษาค้นคว้าต่อไป

#### วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคะแนนสอบวิชาต่าง ๆ 3 วิชา ระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับ การให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว (Three - parameter Logistic Model : 3PL) วิชาที่นำมาเปรียบเทียบ 3 วิชา คือ คณิตศาสตร์ วิทยาศาสตร์ และพละนาฏย โดยทำการเปรียบเทียบในแต่ละวิชาพร้อมทั้งหาความสัมพันธ์ของคะแนนสอบ และเปรียบเทียบค่าความเที่ยงของแบบสอบ จากการตรวจให้คะแนนที่แตกต่างกันในหัวข้อต่อไปนี้

1. เปรียบเทียบมัชฌิมเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ ระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากันตามทฤษฎีมาตรฐานเดิม กับการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีการสนองตอบข้อกระทง
2. หาความสัมพันธ์ของคะแนนสอบระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากัน ตามทฤษฎีมาตรฐานเดิมกับการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีการสนองตอบข้อกระทง
3. เปรียบเทียบค่าความเที่ยงของแบบสอบระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากันตามทฤษฎีมาตรฐานเดิมกับการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีการสนองตอบข้อกระทง

#### สมมติฐานของการวิจัย

จากงานวิจัยของ เต คี ฮู และ ชูศักดิ์ ชัมภลสิทธิ์ (Tse - Chi Hsu and Choosak khampalikit 1982 : 10 - 16) ได้ศึกษาเปรียบเทียบผลของการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method : W) ตามทฤษฎีการสนองตอบข้อกระทง (IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) กับการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) คะแนนมาตรฐาน (Z) และคะแนนความสามารถ ( $\hat{\theta}$ ) โดยศึกษาข้อมูลจากสภาพการสอบคัดเลือกจริง และได้ใช้แบบสอบสองฉบับที่มีลักษณะต่างกันคือ แบบสอบภาษาจีน และแบบสอบประวัติศาสตร์ ซึ่งเป็นแบบสอบที่เขียนขึ้นใหม่และยังไม่มี การทดลองหรือใช้มาก่อนเลย วิธีการเปรียบเทียบคือสมมติว่าถ้าจะรับผู้สมัครสอบเข้าเรียน 10% ของจำนวนผู้สมัครแล้ว การเรียงลำดับตามแบบการให้คะแนนชนิดต่าง ๆ จะให้จำนวนผู้ที่สอบได้ทั้งสองวิธี ได้ก็เปอร์เซ็นต์ โดยคิดจากคะแนนรวมของแบบสอบทั้งสองฉบับ ผลปรากฏว่าในการสอบวัดดังกล่าว สามารถจะนำข้อมูลจากการสอบไปใช้กับทฤษฎีการสนองตอบข้อกระทง (IRT) ได้ และพบว่าเมื่อใช้การให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) กับการสอบคัดเลือกแล้วคุณภาพของข้อกระทงที่เขียนขึ้นมาใหม่นั้นดูเหมือนจะไม่เป็นปัญหาสำหรับการให้คะแนนโดยวิธีนี้ กล่าวคือ การให้คะแนนแบบดังกล่าวจะช่วยปรับปรุงคุณภาพของแบบสอบไปในตัว และในด้านความสัมพันธ์ของคะแนนพบว่าคะแนนจาก

การให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) กับแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) มีความสอดคล้องกันสูงถึง 87% ของจำนวนผู้สมัครสอบ 10% ที่ถูกคัดเลือกโดยคะแนนทั้งสองแบบ

งานวิจัยของแพทเนอิก และทรูป (Patnaik and Traub 1973 : 281 - 285) พบว่าวิธีตรวจให้คะแนนโดยกำหนดคะแนนตัวเลือกถูกในแต่ละข้อให้มีคะแนนต่างกันเมื่อเปรียบเทียบกับวิธี 0 - 1 แล้ว มีค่าความเที่ยงของแบบสอบเป็น .915 และ .881 ตามลำดับ

และจากงานวิจัยของโอลาเกเบต (Olagbade 1971 : 1922 - A) ซึ่งได้ศึกษาผลของการให้คะแนนจากแบบสอบที่ครูสร้างขึ้น 5 วิธีคือ I. วิธีความน่าจะเป็น (Probabilistic) II. วิธีตัดข้อผิดออกโดยมีการตัดคะแนน (elimination - with - penalty) III. วิธีตัดข้อผิดออกโดยไม่มีการตัดคะแนน (elimination - without - penalty) IV. วิธีให้น้ำหนักตัวเลือก (Weighted choice) และ V. วิธีปกติ (conventional) ผลปรากฏว่าคะแนนเฉลี่ยจากวิธีให้คะแนนที่แตกต่างกันมีค่าดังนี้ 127.16 (IV), 109.50 (V), 106.40 (I), 108.82 (III), 92.78 (II) ยกเว้นวิธี II และ III แล้ว พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญ ( $P < .05$ ) สหสัมพันธ์ระหว่างวิธีให้คะแนนแต่ละแบบมีค่าสูงมาก (.90 - .98) โดยวิธี I, II และ III มีสหสัมพันธ์สูงสุด วิธี III และ IV มีสหสัมพันธ์ต่ำสุด.

และจากงานวิจัยของ ชิน ชิง เวิน (Shih-Sung Wen 1975 : 197 - 199) พบว่าสหสัมพันธ์ของคะแนนที่นับจากผลรวมของจำนวนข้อที่ตอบถูก กับคะแนนการทดสอบความมั่นใจมีค่าสูงถึง .97

นอกจากนี้การให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎี การสนองตอบข้อกระทง เป็นเทคนิควิธีที่จะให้คะแนนที่สูงที่สุดและเหมาะสมสำหรับผู้สอบแต่ละคน จากข้อกระทง แต่ละข้อภายใต้ขีดจำกัดของค่าพารามิเตอร์ของข้อกระทง (a, b, c) และค่าระดับความสามารถของผู้สอบ ( $\theta$ ) เอง ด้วยเหตุผลและงานวิจัยดังกล่าว ผู้วิจัยจึงตั้งสมมติฐานของการวิจัยดังนี้

1. มัชฌิม เลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ จากการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical

Test Theory) กับการให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) ภายในแต่ละวิชาน่าจะแตกต่างกัน

2. คะแนนลอบของนักเรียนแต่ละคนระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับการให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) ภายในแต่ละวิชาน่าจะแตกต่างกัน

3. ค่าความเที่ยงของแบบลอบ (Reliability) แต่ละวิชา จากการให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) น่าจะสูงกว่าการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory)

#### ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้มุ่งเปรียบเทียบคะแนนลอบเฉพาะเมื่อให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับการให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว (Three - parameter Logistic Model : 3 PL)

2. การวิจัยครั้งนี้เป็นการวิเคราะห์ข้อมูลที่มีอยู่แล้ว (Re - analysis) คือคะแนนลอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัย ของสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ จากการลอบนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศตามโครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ซึ่งตรวจให้คะแนนแบบให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory)

3. การวิจัยครั้งนี้ใช้ข้อมูลซึ่งสุ่มจากแบบแบ่งชั้นภูมิ (Stratified Random Sampling) จากทุกเขตการศึกษา เขตการศึกษาละ 2.7% ของจำนวนประชากร รวมทั้งสิ้นวิชาละ 3,907 จำนวนเท่านั้น

#### ข้อตกลงเบื้องต้น

1. ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นคะแนนสอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัย ที่มีอยู่แล้วของสำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ จากการสอบนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ระหว่างวันที่ 13 - 14 ธันวาคม พ.ศ. 2526 ตามโครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ตรวจสอบให้คะแนนแบบให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) โดยใช้แบบสอบฉบับเดียวกัน แบบสอบแต่ละฉบับสร้างโดยคณะกรรมการสร้างเครื่องมือแต่ละกลุ่มวิชา และผ่านการพิจารณาจากสำนักงานคณะกรรมการอำนวยการ นักเรียนทุกคนสอบในเวลาเดียวกันภายใต้วิธีดำเนินการสอบเหมือนกัน มีคณะกรรมการวิเคราะห์และประเมินผลทำหน้าที่ดำเนินการวางแผนการสกัดกระทำกับแบบสอบที่ได้จากการสอบ โดยดำเนินงานทั้งก่อนและหลังการสอบเสร็จสิ้นลง มี ดร.วัลลภ กันททรัพย์ เป็นหัวหน้าฝ่ายวิเคราะห์และประมวลผลเป็นผู้รับผิดชอบโดยตลอด การดำเนินการทุกขั้นตอนต้องได้รับความเห็นชอบจากคณะกรรมการอำนวยการ (เยี่ยม โต้บุญเลี้ยง, สัมภาษณ์) ข้อมูลแต่ละวิชาอยู่ในแบบบันทึกข้อมูลแยกเป็นรายวิชา ๆ ละ 1 ม้วน ถือว่าได้รับการตรวจสอบแล้ว และมีความเชื่อถือได้

2. ค่าความเที่ยงของแบบสอบ (Reliability) ตามทฤษฎีการสนองตอบข้อกระทงอยู่ในรูปของ Test information function แต่ในการวิจัยครั้งนี้จะใช้ค่าที่คำนวณจากคะแนนการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีการสนองตอบข้อกระทง โดยสูตรครอนบาค แอลฟา (Cronbach's Alpha)

#### คำจำกัดความที่ใช้ในการวิจัย

การให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory)

หมายถึง การตรวจให้คะแนนข้อกระทงแต่ละข้อโดยกำหนดให้ข้อกระทงทุกข้อมีน้ำหนักคะแนนเท่ากันคือตอบถูกได้ 1 ตอบผิดหรือไม่ตอบได้ 0



การให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตาม ทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT)

หมายถึง การตรวจให้คะแนนข้อกระทงแต่ละข้อโดยกำหนดให้แต่ละข้อ กระทงมีน้ำหนักคะแนนแตกต่างกันตามระดับความสามารถของผู้สอบแต่ละคน (θ) และค่าพารามิเตอร์ของข้อกระทงนั้น 3 ตัว คือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c)

นักเรียน หมายถึง นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ที่เข้าสอบใน โครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ซึ่งได้จากการสุ่มจำนวนวิชาละ 3,907 คน

แบบสอบคณิตศาสตร์ หมายถึง แบบสอบที่สร้างโดยคณะกรรมการสร้างเครื่องมือกลุ่ม คณิตศาสตร์ในโครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ข้อกระทงเป็นคำถามเฉพาะด้านความคิด (Cognitive) แบบ 4 ตัวเลือก รวม 60 ข้อ ใช้เวลาสอบ  $1\frac{1}{2}$  ชั่วโมง

แบบสอบวิทยาศาสตร์ หมายถึง แบบสอบที่สร้างโดยคณะกรรมการสร้างเครื่องมือกลุ่ม วิทยาศาสตร์ในโครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ข้อกระทงเป็นคำถามเฉพาะด้านความคิด (Cognitive) แบบเลือกตอบ 4 ตัวเลือก รวม 60 ข้อ ใช้เวลาสอบ  $1\frac{1}{2}$  ชั่วโมง

แบบสอบพลานามัย หมายถึง แบบสอบที่สร้างโดยคณะกรรมการสร้างเครื่องมือกลุ่ม พลานามัยในโครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ข้อกระทงเป็นคำถามเฉพาะด้านความคิด (Cognitive) แบบเลือกตอบ 4 ตัวเลือก รวม 20 ข้อ ใช้เวลาสอบ  $\frac{1}{2}$  ชั่วโมง

#### ประโยชน์ที่คาดว่าจะได้รับ

เนื่องจากการประยุกต์ใช้ทฤษฎีการสนองตอบข้อกระทง (IRT) ในด้านของการ ตรวจให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ยังไม่เป็นที่รู้จักกันแพร่หลายนักในประเทศไทย และยังไม่มีการวิจัยมาก่อน ผู้วิจัยคาดว่า การวิจัยครั้งนี้จะมีประโยชน์ดังนี้

1. ทำให้ทราบผลของการใช้วิธีตรวจให้คะแนนที่แตกต่างกันว่ามีค่ามัธยฐานเลข-คณิต ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบและคุณภาพด้านความเที่ยงของแบบสอบอย่างไร
2. เป็นแนวทางในการนำไปประยุกต์ใช้เพื่อเพิ่มคุณภาพของแบบสอบและความเชื่อถือได้ของคะแนนสอบ ทำให้การวัดผลมีประสิทธิภาพมากที่สุด
3. เพิ่มพูนความรู้ด้านการวัดผลการศึกษา ซึ่งเป็นประโยชน์ในวงการศึกษา และพัฒนาทฤษฎีวัดผลการศึกษาในประเทศไทยให้ก้าวหน้า
4. เป็นแนวทางในการศึกษาริ้วยต่อไป



ศูนย์วิทยพัทยากร  
จุฬาลงกรณ์มหาวิทยาลัย