บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวัดผลทางการศึกษาและจิตวิทยานั้น สิ่งที่เราต้องการวัดและตรวจสอบส่วนใหญ่เป็นคุณลักษณะภายใน (Latent Traits) ที่ไม่สามารถสังเกตได้โดยตรง เช่น ความสามารถในการคิดคำนวณ ความสามารถในการทำความเข้าใจเนื้อหาต่าง ๆ เป็นต้น การวัดผลจึงจำเป็นต้องอาศัยข้อสอบเป็นสิ่งเร้า เพื่อให้บุคคลแสดงพฤติกรรมตอบสนองต่อข้อสอบเหล่านั้น แล้วจึงนำผลการตอบสนองของบุคคลนั้นมาอนุมานถึงความสามารถที่แท้จริง อันเป็นคุณสมบัติภายในของบุคคลนั้น

ทฤษฎีการวัดผลที่ใช้ทำนายหรืออธิบายความสามารถที่แท้จริงของบุคคลที่ใช้กันอยู่ทั่ว ๆ ไป คือ ทฤษฎีดั้งเดิม (The Classical Test Theory) ซึ่งการประมาณค่าคะแนนจริง (True Score) หรือความสามารถที่แท้จริงตามทฤษฎีนี้ กระทำโดยการนับจำนวนข้อที่ตอบถูกมาใช้ในการประมาณค่า ซึ่งถูกมองว่าไม่เพียงพอต่อการประมาณค่าความสามารถที่แท้จริง และไม่สามารถแก้ปัญหาอีกหลายประการของการวัดผลในปัจจุบัน นักวัดผลจึงพยายามเสนอแนวความคิดใหม่ เพื่อแก้ปัญหาเหล่านั้น ในระยะปัจจุบันมีทฤษฎีที่น่าสนใจที่พัฒนาขึ้นทฤษฎีหนึ่ง คือทฤษฎีที่ใช้โมเดล (Model) ทางคณิตศาสตร์สำหรับแสดงความสัมพันธ์ ระหว่างความสามารถที่แท้จริงกับพฤติกรรมการตอบข้อสอบของแต่ละบุคคล ทฤษฎีนี้มีชื่อเรียกหลายชื่อ คือ ทฤษฎีคุณลักษณะแฝง(Latent Trait Theory) หรือ ทฤษฎีการตอบข้อคำถาม (Item Response Theory) หรือ ทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) (ผจงจิตร, 2525)

ทฤษฎีคุณลักษณะแฝงให้ประโยชน์อย่างมากแก่วงการวัดผลการศึกษา โดยเฉพาะอย่างยิ่งปัญหาการจัดข้อสอบให้สอดคล้องกับความสามารถของบุคคล ซึ่งทฤษฎีการวัดผลแบบดั้งเดิมไม่สามารถแก้ปัญหานี้ได้ กล่าวคือ การทดสอบแบบดั้งเดิมนั้นแบบทดสอบจะประกอบด้วยกลุ่มของข้อสอบที่กำหนดจำนวนข้อสอบไว้แน่นอน มีความยากโดยเฉลี่ยอยู่ในระดับปานกลาง ซึ่งเหมาะสมกับผู้สอบที่มีความสามารถระดับปานกลางเท่านั้น แต่ในสภาพการทดสอบจริง ๆ แล้ว ผู้สอบแต่ละคนจะมีความสามารถในระดับที่แตกต่างกัน นั่นคือ ไม่ว่าผู้สอบจะมีความสามารถในระดับใด ก็จะต้องทำแบบทดสอบฉบับเดียวกันหรือทำข้อสอบเหมือนกันหมดทุกคน ดังนั้นการทดสอบแบบดั้งเดิมจะมีความเหมาะสมในกรณีที่ผู้สอบมีความสามารถระดับปานกลาง หรืออยู่ในช่วงใกล้ค่าเฉลี่ยของกลุ่มผู้สอบ ส่วนผู้สอบที่มีความสามารถสูง หรือต่ำกว่าระดับความสามารถเฉลี่ยของกลุ่มผู้สอบมาก ๆ จะได้ข้อมูลเกี่ยวกับตัวผู้สอบ

ต่ำ (Vale and Weiss, 1975) ในกรณีเช่นนี้ ถ้าผู้สอบได้ข้อสอบที่มีความยากสูงกว่าระดับความสามารถของตนเองแล้ว ผู้สอบมีแนวโน้มที่จะตอบข้อสอบด้วยพฤติกรรมอื่น ๆ นอกเหนือจากความรู้ความสามารถที่มีอยู่ เช่น ใช้การเดาในการตอบข้อสอบ และถ้าผู้สอบคนใดได้ตอบข้อสอบที่มีความยากต่ำกว่าระดับความสามารถของตนเองแล้ว ข้อสอบข้อนั้นจะไม่ยั่วยุให้ผู้สอบได้ใช้ความสามารถในระดับสูงสุดของเขาหรือทำให้เกิดความเบื่อหน่าย ซึ่งปัจจัยดังกล่าวนี้จะส่งผลกระทบต่อการตอบข้อสอบของผู้สอบทั้งสองกลุ่มที่ไม่ได้ให้ผลในระดับสูงสุด ทำให้ความเชื่อมั่นของคะแนนผลการสอบจากผู้สอบทั้งสองกลุ่มนี้ ต่ำกว่าที่ควรจะเป็น (Weiss, 1974)

ปัจจุบันปัญหานี้สามารถแก้ไขได้โดยใช้การทดสอบแบบเทเลอร์ (Tailored Testing) ซึ่งเป็นผลมาจากการประยุกต์ใช้ทฤษฎีคุณลักษณะแฝงอย่างได้ผล (Ruder, 1983) การทดสอบแบบเทเลอร์ มุ่งที่จะเลือกข้อสอบให้มีระดับความยากที่เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน การสอบในแต่ละครั้งผู้สอบแต่ละคนอาจไม่จำเป็นต้องทำข้อสอบเหมือนกันหมดทุกข้อทั้งฉบับ แต่ทั้งนี้จำนวนข้อที่แต่ละคนทำ ขึ้นอยู่กับผลการตอบข้อสอบแต่ละข้อในแบบทดสอบของแต่ละคน (Weiss and Kingsbury, 1984) การทดสอบแบบเทเลอร์โดยทั่วไป ผู้สอบจะเริ่มทำข้อสอบข้อที่มีความยากระดับปานกลาง ถ้าผู้สอบตอบข้อสอบถูกข้อสอบข้อต่อไปที่จะได้ทำก็จะมีความยากมากขึ้น แต่ถ้าตอบข้อสอบผิดข้อต่อไปที่จะได้ทำก็จะง่ายลง และการทำข้อสอบข้อต่อ ๆ ไปก็จะปฏิบัติเช่นเดียวกันนี้จนสิ้นสุดการสอบ (Green and Others, 1984) และจากการศึกษาของ Betz and Weiss ในปี ค.ศ. 1973 (อ้างถึงใน Larkin and Weiss, 1975) พบว่า การทดสอบแบบเทเลอร์จะใช้ข้อสอบน้อยกว่าการทดสอบแบบดั้งเดิม จึงใช้เวลาในการทำแบบทดสอบน้อยลง ในขณะที่ความเชื่อมั่น (Reliability) และความเที่ยงตรง (Validity) คงเดิมในทุกระดับความสามารถของผู้สอบตลอดจนช่วยกระตุ้นให้ผู้สอบได้ใช้ความรู้ และความสามารถในการทำแบบทดสอบอย่างเต็มที่อีกด้วย นอกจากนี้ยังพบว่า การทดสอบแบบเทเลอร์ที่ดำเนินการสอบด้วยคอมพิวเตอร์ มีประโยชน์อีกหลายประการ ดังนี้ (Urry, 1977)

1. การดำเนินการสอบมีมาตรฐาน จึงขจัดอคติที่อาจเกิดจากผู้ดำเนินการสอบ
2. ข้อสอบไม่รั่ว เพราะไม่มีการพิมพ์ข้อสอบลงบนกระดาษ และผู้สอบแต่ละคนก็ทำข้อสอบที่ได้รับต่างกัน
3. ช่วยปรับปรุงการวางโปรแกรมการสอบสำหรับผู้สมัครสอบเป็นจำนวนมาก ๆ
4. ช่วยปรับปรุงความเชื่อมั่นของการวัดให้มีประสิทธิภาพสูงขึ้น

การทดสอบแบบเทเลอร์มีวิธีการสร้างและการดำเนินการสอบอยู่ 2 ประเภทใหญ่ ๆ ดังแสดงไว้ในภาพที่ 1 คือ วิธีการสองขั้น (Two-Stage Strategies) และ วิธีการหลายขั้น (Multi-Stage Strategies)(Weiss,1974) วิธีการหลายขั้นประกอบด้วยแบบจำลอง 2 รูปแบบ คือ แบบจำลองการแยกทางคงที่ (Fixed Branching Models) และแบบจำลองการแยกทางแปรผัน (Variable Branching Models) แบบทดสอบที่ใช้แบบจำลองการแยกทางคงที่ได้แก่แบบทดสอบปิรามิด (Pyramidal Test) แบบทดสอบเฟล็กซิเลเวล (Flexilevel Test) และแบบทดสอบปรับระดับขั้น (Stradaptive Test) ส่วนแบบจำลองการแยกทางแปรผันแบ่งออกเป็น 2 วิธีการ คือ วิธีการของเบย์ (Bayesian Strategies) และ วิธีการความเป็นไปได้สูงสุด (Maximum Likelihood Strategies) งานวิจัยเกี่ยวกับการทดสอบแบบเทเลอร์ในต่างประเทศที่ผ่านมา ส่วนใหญ่จะศึกษาเกี่ยวกับวิธีการการทดสอบแบบแยกทางคงที่ (Weiss and Betz, 1973)

แบบทดสอบปิรามิด หรือแบบทดสอบโครงสร้างต้นไม้ (Tree-Structure Test) เท่าที่ศึกษากันมีอยู่ 5 ลักษณะคือ การใช้ขนาดขั้นคงที่ (Constant Step Size) การใช้ขนาดขั้นแปรผัน (Variable Step Size) ปิรามิดช้างตัด (Truncated Pyramids) รูปจำลองที่มีข้อสอบหลายข้อในแต่ละขั้น (Multiple - Item Models) และการให้ค่าน้ำหนักตัวเลือกเพื่อแยกทาง (Differential Response Option Branching) (Weiss, 1974) แบบทดสอบปิรามิดมีคุณลักษณะเด่นหลายประการดังนี้

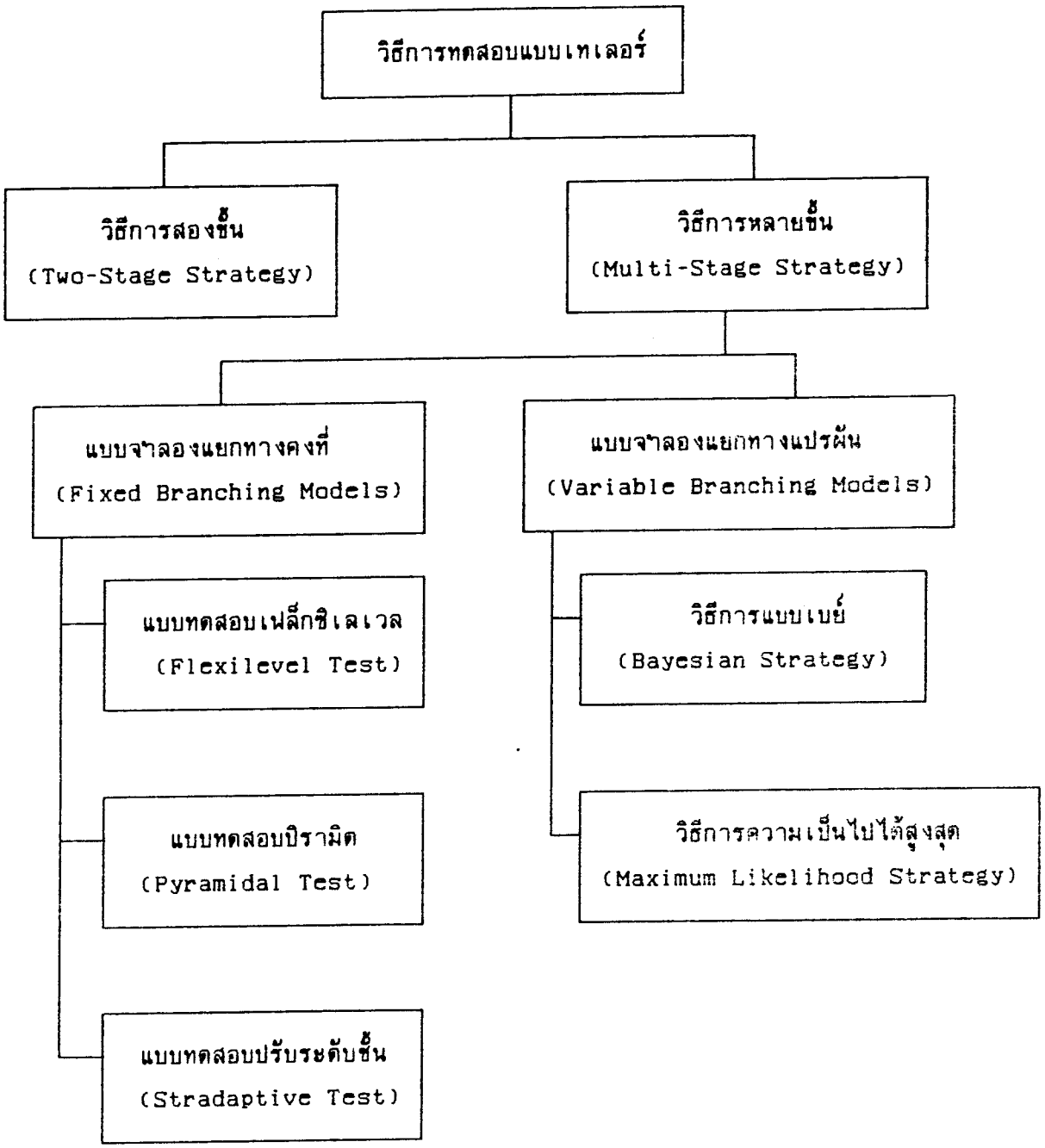
1. จัดระดับความยากของข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคนได้ ทำให้ผลการสอบมีความเชื่อมั่นมากขึ้น

2. จำนวนข้อสอบในแบบทดสอบที่ผู้สอบจะได้ตอบมีจำนวนลดลง เมื่อเทียบกับแบบทดสอบดั้งเดิม ซึ่งมีผลทำให้ใช้เวลาในการสอบลดลงเมื่อเทียบกับเวลาที่ใช้ในการสอบของแบบทดสอบดั้งเดิม

3. วัดความสามารถของผู้สอบในขอบเขตที่กว้างกว่าแบบทดสอบชนิดสองขั้นด้วยจำนวนข้อสอบที่น้อยกว่า

4. ดำเนินการสอบได้ทั้งใช้การเขียนตอบและใช้คอมพิวเตอร์ช่วย

จากวิธีการทดสอบแบบเทเลอร์ ที่ต้องมีการปรับระดับความยากของข้อสอบ หรือเลือกข้อสอบที่มีระดับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน ดังนั้นแบบทดสอบเทเลอร์จึงมีข้อจำกัดในการสร้างและวิธีดำเนินการสอบ กล่าวคือ การสร้างแบบทดสอบเทเลอร์จะต้องใช้ข้อสอบที่ใช้วัดความสามารถด้านเดียวกัน (Unidimensionality) จำนวนมากหรือต้องมีกลุ่มข้อสอบขนาด



ภาพที่ 1 แสดงรูปแบบของแบบทดสอบเทเลอร์

ใหญ่ และข้อสอบต้องผ่านการทดลองใช้เพื่อหาค่านารามิเตอร์ของข้อสอบ (Item Parameters) มาก่อน เพื่อจะได้ค่าความยากและค่าอำนาจจำแนกที่คงที่ไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ ส่วนวิธีดำเนินการสอบโดยการเขียนตอบนั้น ผู้สอบมักจะไม่เข้าใจ หรือเกิดปัญหาและความยุ่งยากในการปฏิบัติตามคำชี้แจงของแบบทดสอบ ซึ่งมีผลทำให้ผู้สอบเกิดความกังวลใจในการสอบ (ผดุง,2530) การใช้วิธีดำเนินการสอบโดยการเขียนตอบจึงเป็นปัญหาของการทดสอบแบบเทเลอร์ ปัญหานี้น่าจะสามารถแก้ไขได้โดยการนำคอมพิวเตอร์เข้ามาช่วยในการดำเนินการสอบ ซึ่งมีขั้นตอนในการดำเนินการดังนี้

1. ให้ผู้สอบนั่งหน้าเครื่องคอมพิวเตอร์ที่มีโปรแกรมของแบบทดสอบเทเลอร์
2. ข้อสอบจะปรากฏบนจอภาพ โดยให้ผู้สอบตอบทีละข้อ
3. ผู้สอบตอบข้อสอบโดยการกดคำตอบลงบนแป้นพิมพ์
4. ถ้าคำตอบถูก โปรแกรมดำเนินการสอบจะเรียกข้อสอบข้อต่อไปที่มีความยากสูงขึ้นให้ปรากฏที่หน้าจอ แต่ถ้าคำตอบผิดข้อสอบข้อต่อไปที่จะได้รับก็จะมีความยากลดลง
5. เมื่อถึงเกณฑ์หยุดการสอบ ซึ่งขึ้นอยู่กับรูปแบบของแบบทดสอบเทเลอร์ที่ใช้ คอมพิวเตอร์จะยุติการดำเนินการสอบและแจ้งผลการสอบให้ผู้สอบทราบทันที

จากวิธีการดังกล่าว โปรแกรมคอมพิวเตอร์จะปรับการประมาณค่าความสามารถของผู้สอบไปเรื่อย ๆ ตามค่าระดับความยากของข้อสอบที่ทำ และค่าความสามารถที่ประมาณได้นี้จะมีความเชื่อมั่นสูงขึ้นเรื่อยๆ ตามจำนวนครั้งของการปรับ การที่จะใช้คอมพิวเตอร์ดำเนินการสอบแบบทดสอบเทเลอร์ในลักษณะนี้ได้นั้น จำเป็นต้องมีการพัฒนาโปรแกรมคอมพิวเตอร์ ให้เหมาะสมกับแบบทดสอบเทเลอร์รูปแบบต่าง ๆ ดังนั้นจึงน่าจะมีการนำเทคโนโลยีทางด้านคอมพิวเตอร์มาประยุกต์ใช้ทางการวัดผลทางการศึกษา เนื่องสร้างสรรค์แนวทางใหม่ของการวัดผลที่มีประสิทธิภาพได้อย่างกว้างขวางยิ่งขึ้น ซึ่งเป็นสิ่งที่มีความเป็นไปได้สูง ทั้งนี้เนราะในปัจจุบันการศึกษาระดับมัธยมศึกษามีแนวโน้มที่โรงเรียนต่าง ๆ จะมีเครื่องคอมพิวเตอร์ใช้กันมากขึ้น เนราะในหลักสูตรใหม่ของกระทรวงศึกษาธิการได้เพิ่มรายวิชา ค 031 ความรู้เบื้องต้นเกี่ยวกับคอมพิวเตอร์ และ รายวิชา ค 032 การเขียนโปรแกรมภาษาเบสิก เป็นวิชาเลือกในหลักสูตรวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาตอนปลาย (กระทรวงศึกษาธิการ,2528) ดังนั้นจึงถือเป็นโอกาสดีที่การทดสอบแบบเทเลอร์จะได้อาศัยระบบคอมพิวเตอร์มาใช้ในการดำเนินการสอบเพื่อพัฒนาการทดสอบให้มีประสิทธิภาพสูงขึ้น และทำให้นักเรียนระดับมัธยมศึกษามีความคุ้นเคยกับการใช้คอมพิวเตอร์มากขึ้น

จากลักษณะเด่นของทฤษฎีคุณลักษณะแฝง การทดสอบแบบเทเลอร์ และการใช้คอมพิวเตอร์ช่วยในการดำเนินการสอบดังกล่าว ผู้วิจัยมีความสนใจที่จะศึกษาประสิทธิภาพของการดำเนินการสอบแบบปิรามิดด้วยคอมพิวเตอร์ สำหรับทดสอบวัดผลสัมฤทธิ์ทางการเรียนในชั้นเรียนในสภาพการณ์จริง

เนราะการศึกษาในแนวนี้ยังไม่เป็นที่แพร่หลายในแวดวงการทดสอบของประเทศไทย นอกจากนี้ผลการศึกษาอาจจะช่วยให้เห็นแนวทางในการสร้าง วิธีดำเนินการสอบ และประเมินประสิทธิภาพแบบทดสอบเทเลอร์ที่ดำเนินการสอบด้วยคอมพิวเตอร์ ที่มีประสิทธิภาพอีกรูปแบบหนึ่ง

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ดังต่อไปนี้

1.2.1 เพื่อสร้างแบบทดสอบปิรามิดวิชาเคมีระดับชั้นมัธยมศึกษาปีที่ 5 เรื่อง อัตราการเกิดปฏิกริยาเคมี

1.2.2 เพื่อพัฒนาโปรแกรมคอมพิวเตอร์ควบคุมการดำเนินการสอบแบบทดสอบปิรามิดที่สร้างขึ้น

1.2.3 เพื่อประเมินประสิทธิภาพของแบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ที่สร้างขึ้น

1.2.4 เพื่อศึกษาความสัมพันธ์ระหว่างรูปแบบของแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิม กับความสามารถทางการเรียนในวิชาเคมี

1.2.5 เพื่อประเมินประสิทธิภาพของโปรแกรมคอมพิวเตอร์ควบคุมการดำเนินการสอบ แบบทดสอบปิรามิดที่พัฒนาขึ้น

1.3 สมมติฐานการวิจัย

1.3.1 แบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ทั้งสองฉบับ มีค่าความเชื่อมั่นสูงกว่าแบบทดสอบดั้งเดิมที่ระดับนัยสำคัญทางสถิติ .05

1.3.2 แบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ทั้งสองฉบับ มีประสิทธิภาพด้านความตรงเชิงสัมพันธ์กับเกณฑ์ที่ระดับนัยสำคัญทางสถิติ .05

1.3.3 แบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ทั้งสองฉบับ มีประสิทธิภาพด้านความตรงเชิงสัมพันธ์กับเกณฑ์สูงกว่าแบบทดสอบชนิดดั้งเดิม ที่ระดับนัยสำคัญทางสถิติ .05

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้แบบทดสอบวิชาเคมีระดับชั้นมัธยมศึกษาปีที่ 5 เรื่อง อัตราการเกิดปฏิกริยาเคมี 3 ฉบับ ประกอบด้วย แบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ 2 ฉบับ และแบบทดสอบดั้งเดิม 1 ฉบับที่สามารถใช้ในการเรียนการสอนในชั้นเรียนได้

1.4.2 ได้โปรแกรมคอมพิวเตอร์ควบคุมการดำเนินการสอบแบบทดสอบปิรามิด 2 โปรแกรม

1.4.3 การวิจัยครั้งนี้เป็นแนวทางในการศึกษาประสิทธิภาพของการทดสอบแบบปิรามิดด้วยคอมพิวเตอร์ให้กว้างขวางยิ่งขึ้น

1.4.4 การวิจัยครั้งนี้ทำให้ได้ข้อมูลในการพิจารณาปรับปรุงและพัฒนารูปแบบ และวิธีการดำเนินการสอบแบบทดสอบเทเลอร์ด้วยคอมพิวเตอร์ให้มีประสิทธิภาพมากขึ้น

1.5 นิยามปฏิบัติการ

1.5.1 ข้อสอบ หมายถึง ข้อคำถามเนื้อหาวิชาเคมีชั้นมัธยมศึกษาปีที่ 5 เรื่องอัตราการเกิดปฏิกริยาเคมี ซึ่งคำถามแต่ละข้อเป็นแบบเลือกตอบชนิด 4 ตัวเลือก

1.5.2 แบบทดสอบ หมายถึงกลุ่มของข้อสอบเนื้อหาวิชาเคมีชั้นมัธยมศึกษาปีที่ 5 เรื่องอัตราการเกิดปฏิกริยาเคมี ที่เป็นแบบเลือกตอบชนิด 4 ตัวเลือก

1.5.3 แบบทดสอบปิรามิด (Pyramidal Test) หมายถึง แบบทดสอบที่จัดเรียงข้อสอบตามแบบจำลองรูปปิรามิด ซึ่งยึดถือค่าความยากของข้อสอบตามโครงสร้างที่ฐานและระดับชั้นของรูปปิรามิดที่บันทึกลงในแผ่นจานแม่เหล็ก (Floppy Disk or Diskette) ของเครื่องคอมพิวเตอร์ ดำเนินการสอบโดยให้ผู้สอบนั่งหน้าเครื่องคอมพิวเตอร์ แจ้งเลขประจำตัวโดยกดลงบนแป้นพิมพ์ จากนั้นเครื่องคอมพิวเตอร์ก็จะเริ่มแสดงข้อสอบข้อแรกทางจอภาพให้นักเรียนทำ เมื่อผู้สอบตอบข้อสอบโดยการกดลงบนแป้นพิมพ์แล้ว เครื่องคอมพิวเตอร์ก็จะนำผลการตอบนั้นไปพิจารณาเลือกข้อสอบข้อต่อไปให้ผู้สอบทำตามหลักเกณฑ์ของแบบทดสอบปิรามิดที่ใช้ซึ่งแบ่งออกเป็น 2 ฉบับ คือ

1.5.3.1 ฉบับที่ 1 เป็นแบบทดสอบปิรามิดชนิด 8 ชั้นที่ใช้ขนาดขั้นคงที่ (Constant Step Size) และกำหนดแนวทางการตอบแบบเพิ่ม 1/ลด 1 (Up-One/Down-One Branching Rule) ซึ่งหมายถึง แบบทดสอบที่สร้างขึ้นตามแบบจำลองที่มีจำนวนข้อสอบในแต่ละชั้นเท่ากับลำดับที่ของชั้น คือ ในชั้นที่ 1 จะมีข้อสอบ 1 ข้อ ในชั้นที่ 8 จะมีข้อสอบ 8 ข้อ ดังนั้นแบบทดสอบฉบับนี้มีจำนวนข้อสอบทั้งหมด 36 ข้อ และค่าความแตกต่างระหว่างค่าความยากของข้อสอบในแต่ละชั้นที่อยู่ติดกัน จะมีค่าเท่ากันหรือคงที่ นักเรียนต้องตอบข้อสอบฉบับนี้ชั้นละ 1 ข้อ รวมจำนวนข้อสอบที่นักเรียนจะต้องตอบ 8 ข้อ

1.5.3.2 ฉบับที่ 2 เป็นแบบทดสอบปิรามิดชนิด 5 ชั้น แบบที่มีข้อสอบชุดละ 3 ข้อในแต่ละชั้น ที่ใช้ขนาดขั้นคงที่ และกำหนดแนวทางการตอบแบบ เพิ่ม1 / ลด 1 การกำหนดข้อสอบให้ทำในขั้นต่อไปใช้คะแนนที่ได้จากการตอบถูกในขั้นที่มาก่อน ซึ่งหมายถึง แบบทดสอบที่สร้างขึ้นตามแบบจำลองที่มีจำนวนข้อสอบชุดละ 3 ข้อในแต่ละชั้นนี้ ข้อสอบในแต่ละชุด (3 ข้อ) จะต้องมีความยากใกล้เคียงกัน โดยที่ข้อสอบ 3 ข้อแรก จะเป็นข้อสอบที่มีความยากระดับปานกลาง ผู้สอบต้องตอบทั้ง 3 ข้อและแต่ละข้อจะให้คะแนนแบบ 1-0 ผู้สอบที่ได้คะแนน 0 หรือ 1 จะถูกแยก

ทางให้ไปตอบชุดข้อสอบในขั้นที่ 2 ที่มีความยากลดลง ส่วนผู้ที่ตอบได้คะแนน 2-3 ในขั้นที่ 1 จะถูกแยกทางให้ไปตอบชุดข้อสอบในขั้นที่ 2 ที่มีความยากมากขึ้น แบบทดสอบปิรามิดชนิด 5 ขั้นที่มีข้อสอบชุดละ 3 ข้อในแต่ละขั้นนี้จะมีข้อสอบทั้งหมด 45 ข้อ ซึ่งผู้สอบแต่ละคนจะได้ตอบข้อสอบทั้งสิ้น 15 ข้อ

1.5.4 แบบทดสอบดั้งเดิม(Conventional Test) หมายถึงแบบทดสอบที่จัดเรียงข้อสอบตามลำดับเนื้อหาวิชา และในแต่ละเนื้อหาวิชาเรียงข้อสอบจากง่ายไปยาก ประกอบด้วยข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 30 ข้อ ที่มีช่วงระดับความยากของข้อสอบเท่ากับช่วงระดับความยากของข้อสอบในแบบทดสอบรูปปิรามิดในฉบับที่ 1 และ ฉบับที่ 2 ใช้วัดผลสัมฤทธิ์ทางการเรียนรายวิชา ว.033 เคมี ระดับชั้นมัธยมศึกษาปีที่ 5 เรื่องอัตราการเกิดปฏิกริยาเคมี

1.5.5 แบบทดสอบที่มีมิติเดียว(Unidimentional Test) หมายถึง แบบทดสอบที่สร้างขึ้นเพื่อวัดความสามารถหรือคุณลักษณะเดียว หรือมีความเป็นเอกพันธ์กันสูงระหว่างข้อสอบในแบบทดสอบนั้น การวิจัยนี้หาโดยการวิเคราะห์องค์ประกอบด้วยโปรแกรมสำเร็จรูป $SPSS^X$

1.5.6 ค่าความยากของข้อสอบ หมายถึง ค่าความสามารถตรงจุดเปลี่ยนโค้งคุณลักษณะของข้อสอบแต่ละข้อ(Item Characteristic Curve)ซึ่งเป็นระดับความสามารถที่จะตอบข้อสอบนั้นได้ถูกต้องเมื่อไม่มีการเดาด้วยความน่าจะเป็น .50 (Lord, 1980) การวิจัยครั้งนี้หาโดยใช้ทฤษฎีคุณลักษณะแฝงที่ใช้นารามิเตอร์ 3 ตัว ด้วยโปรแกรมสำเร็จรูปโลจิส V (LOGIST V) และคัดเลือกข้อสอบที่มีค่าความยากที่เหมาะสมกับรูปแบบของแบบทดสอบรูปปิรามิดที่ต้องการ

1.5.7 อำนาจจำแนกของข้อสอบ หมายถึงค่าที่เป็นสัดส่วนผกผันกับความชันของโค้งหรือความชันของความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อนั้นถูก ณ จุดเปลี่ยนโค้ง(Lord, 1980) การวิจัยครั้งนี้ประมาณค่าตามทฤษฎีคุณลักษณะแฝงที่ใช้พารามิเตอร์ 3 ตัว ด้วยโปรแกรมสำเร็จรูปโลจิส V และคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ .30 ขึ้นไป

1.5.8 สัมประสิทธิ์การเดา หมายถึงความน่าจะเป็นที่บุคคลหนึ่งซึ่งปราศจากความสามารถจะตอบข้อสอบข้อนั้นได้ถูกต้อง(Lord, 1980) การวิจัยครั้งนี้ประมาณค่าตามตามทฤษฎีคุณลักษณะแฝงที่ใช้พารามิเตอร์ 3 ตัว ด้วยโปรแกรมโลจิส V และคัดเลือกข้อสอบข้อที่มีค่าสัมประสิทธิ์การเดาน้อยกว่า .30

1.5.9 ระดับคะแนนวิชาเคมี หมายถึง ระดับคะแนนรายวิชา ว.033 เคมี ที่กลุ่มตัวอย่างได้รับหลังจากที่อาจารย์ประจำวิชาตัดสินผลการเรียนในภาคเรียนที่ 1 ปีการศึกษา 2531

1.5.10 แบบสอบถามความคิดเห็นของนักเรียนที่มีต่อแบบทดสอบปิรามิด ที่ดำเนินการสอบด้วยคอมพิวเตอร์ หมายถึง แบบสอบถามที่ผู้วิจัยสร้างขึ้น เพื่อใช้ตรวจสอบความคิดเห็นของนักเรียนที่มีต่อแบบทดสอบปิรามิดที่ใช้ในการวิจัย เป็นแบบสอบถามความคิดเห็นชนิด 5 ระดับ จำนวน 20 ข้อความ

1.5.11 ค่าสัมประสิทธิ์ความตรงเชิงสัมพันธ์กับเกณฑ์ หมายถึงค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) ของคะแนนผลการสอบที่ได้จากแบบทดสอบปิรามิดแต่ละฉบับกับระดับคะแนนวิชาเคมีซึ่งใช้เป็นเกณฑ์

1.5.12 แบบทดสอบมีประสิทธิภาพด้านความตรงเชิงสัมพันธ์กับเกณฑ์ หมายถึง แบบทดสอบที่มีค่าสัมประสิทธิ์ความตรงเชิงสัมพันธ์กับเกณฑ์ที่แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแบบทดสอบที่ถือว่า มีประสิทธิภาพด้านความตรงเชิงสัมพันธ์กับเกณฑ์สูงกว่าแบบทดสอบฉบับอื่น ก็ต่อเมื่อแบบทดสอบฉบับนั้น มีค่าสัมประสิทธิ์ความตรงเชิงสัมพันธ์กับเกณฑ์สูงกว่าแบบทดสอบฉบับอื่นอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.5.13 ความเชื่อมั่น (Reliability) ของแบบทดสอบ หมายถึง คุณสมบัติอย่างหนึ่งของแบบทดสอบที่เมื่อทำการวัดซ้ำแล้วจะให้ผลการวัดเหมือนเดิมหรือใกล้เคียงกัน ซึ่งอาจนิจารณาได้จากค่าสหสัมพันธ์ระหว่างคะแนนจากการทดสอบ 2 ครั้งด้วยแบบทดสอบฉบับเดิม และในที่นี้ค่าความเชื่อมั่นของแบบทดสอบจะหมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของคะแนนความสามารถของผู้สอบครั้งที่ 1 กับคะแนนความสามารถของผู้สอบครั้งที่ 2 ซึ่งมีช่วงเวลาห่างกันประมาณ 2 สัปดาห์ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 1.00 หมายความว่าแบบทดสอบมีค่าความเชื่อมั่นสูง

1.5.14 ความคิดเห็นต่อแบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ อยู่ในเกณฑ์ดี หรือปานกลาง หรือ ต่ำ หมายถึง การที่นักเรียนตอบข้อความในแบบสอบถามความคิดเห็นของนักเรียนที่มีต่อแบบทดสอบปิรามิด ที่ดำเนินการสอบด้วยคอมพิวเตอร์ มีค่าเฉลี่ยของระดับความคิดเห็นตั้งแต่ 3.50ขึ้นไป หรืออยู่ระหว่าง 2.50 ถึง 3.49 หรือต่ำกว่า 2.50 ลงไปตามลำดับ

1.5.15 ประสิทธิภาพเชิงปฏิบัติการ หมายถึง ประสิทธิภาพของแบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ในแง่ของการนำไปใช้ เมื่อนักเรียนแสดงความคิดเห็นอยู่ในเกณฑ์ดี

1.5.16 ความเชื่อมั่นชนิดความคงที่ภายในของแบบสอบถาม หมายถึงความคงที่ภายในของข้อความซึ่งข้อความแต่ละข้อวัดในสิ่งเดียวกันกับข้อความข้ออื่น ๆ ทุกข้อโดยปราศจากความคลาดเคลื่อน ในการวิจัยครั้งนี้ใช้ค่าสัมประสิทธิ์แอลฟา(Coefficient Alpha) ในการหาค่าความเชื่อมั่นชนิดความคงที่ภายในของแบบสอบถาม

1.6.ข้อตกลงเบื้องต้น

1.6.1 ข้อสอบที่ใช้ในการศึกษาครั้งนี้ สร้างจากตารางวิเคราะห์จุดประสงค์การเรียนรู้และลักษณะเฉพาะของข้อสอบ (Item Specification) ที่จำกัดขอบเขตเฉพาะเรื่อง อัตราการเกิดปฏิกริยาเคมี จึงถือว่าข้อสอบทุกข้อวัดผลสัมฤทธิ์ทางการเรียนเพียงด้านเดียว (Unidimensionality)

1.6.2 โอกาสที่ผู้สอบจะตอบข้อสอบข้อใดข้อหนึ่งถูกไม่ขึ้นอยู่กับการตอบข้อสอบข้ออื่น ๆ และการจัดเรียงลำดับข้อสอบในแบบทดสอบจะไม่ส่งผลกระทบต่อคะแนนผลการสอบ (Local Independence)

1.6.3 แบบทดสอบทั้ง 3 ฉบับมีความเป็นอิสระกันของข้อสอบ กล่าวคือไม่มีข้อสอบที่เหมือนกัน ตลอดจนการให้คะแนนก็ไม่มีความเกี่ยวข้องกัน

1.6.4 การตอบข้อสอบในแบบทดสอบของนักเรียน เป็นการตอบตามความสามารถที่แท้จริง

1.6.5 คะแนนผลการสอบของนักเรียน ที่ได้จากแบบทดสอบปิรามิด แบบทดสอบดั้งเดิม และคะแนนจากแบบสอบถามความคิดเห็นต่างก็มีการกระจายในลักษณะที่เป็นโค้งปกติ

1.6.6 การตอบแบบสอบถามความคิดเห็นที่มีต่อแบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ของนักเรียน เป็นการแสดงความรู้สึกที่แท้จริงของนักเรียนแต่ละคน

1.7. ขอบเขตของการวิจัย

1.7.1 กลุ่มตัวอย่าง

1.7.1.1 กลุ่มตัวอย่างสำหรับการสร้างแบบทดสอบปิรามิด ได้จากการสุ่มจากประชากรที่เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 5 ในโรงเรียนระดับมัธยมศึกษาจังหวัดขอนแก่น ที่ศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2531 จำนวน 7 โรงเรียน คือ โรงเรียนขอนแก่นวิทยายน โรงเรียนกัลยาณวัตร โรงเรียนแก่นนครวิทยาลัย โรงเรียนนครขอนแก่น โรงเรียนน้ำพองศึกษา โรงเรียนบ้านไผ่ และโรงเรียนสาธิตมอดินแดง

1.7.1.2 กลุ่มตัวอย่างสำหรับการประเมินประสิทธิภาพ แบบทดสอบปิรามิดที่ดำเนินการสอบด้วยคอมพิวเตอร์ ได้จากการสุ่มจากประชากรที่เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 5 โรงเรียนสาธิตคณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น ที่ศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2531

1.7.2 การวิจัยครั้งนี้มุ่งศึกษาแบบทดสอบปิรามิด 2 รูปแบบ คือ แบบทดสอบปิรามิดชนิด 8 ชั้น ที่มีขนาดชั้นคงที่ ภายใต้กฎการแยกทางแบบเริ่ม 1/ลด 1 และ แบบทดสอบปิรามิดชนิด 5 ชั้นที่นักเรียนต้องทำข้อสอบ 3 ข้อในแต่ละชั้น ซึ่งดำเนินการสอบด้วยคอมพิวเตอร์

1.7.3 เนื้อหาวิชาที่นำมาสร้างข้อสอบคือ เรื่องอัตราการเกิดปฏิกริยาเคมีซึ่งอยู่ในรายวิชา ว.033 วิชาเคมี ระดับชั้นมัธยมศึกษาปีที่ 5

1.7.4 เกณฑ์ที่ใช้ประเมินแบบทดสอบปิรามิด คือ ระดับคะแนนวิชาเคมี จากรายวิชา ว.033 ที่กลุ่มตัวอย่างลงทะเบียนเรียนในภาคเรียนที่ 1 ปีการศึกษา 2531 ซึ่งอาจารย์ประจำวิชาเป็นผู้ตัดสินผลการเรียนโดยให้เป็นระดับคะแนนที่มีค่าเป็น 4,3,2,1 และ 0

1.7.5 การนิจารณาค่าความเชื่อมั่นของแบบทดสอบ ใช้เกณฑ์ดังนี้ แบบทดสอบปิรามิดฉบับที่ 1 ซึ่งนักเรียนได้ทำข้อสอบ 8 ข้อต้องมีค่าความเชื่อมั่นไม่น้อยกว่า .50 แบบทดสอบปิรามิดฉบับที่ 2 ซึ่งนักเรียนได้ทำข้อสอบ 15 ข้อต้องมีค่าความเชื่อมั่นไม่น้อยกว่า .60 และแบบทดสอบดั้งเดิม ซึ่งนักเรียนได้ทำข้อสอบ 30 ข้อต้องมีค่าความเชื่อมั่นไม่น้อยกว่า .70

1.7.6 การประเมินประสิทธิภาพเชิงปฏิบัติการของแบบทดสอบปิรามิด ที่ดำเนินการสอบด้วยคอมพิวเตอร์ ประเมินจากความคิดเห็นของนักเรียน โดยใช้แบบสอบถามความคิดเห็นของนักเรียน