

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) วิเคราะห์ค่าความยากของข้อสอบและค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ข้อสอบแบบพหุระดับ ตามเงื่อนไขของจำนวนผู้สอบและความยาวของแบบสอบ 2) วิเคราะห์ความสัมพันธ์ระหว่างความยากของข้อสอบและความสามารถของผู้สอบที่ได้จากการวิเคราะห์ข้อสอบแบบพหุระดับกับความยากของข้อสอบและความสามารถของผู้สอบที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ 3) วิเคราะห์ความแปรปรวนของค่าความยากของข้อสอบระหว่างโรงเรียนภายในสังกัดเดียวกัน และระหว่างสังกัด และศึกษาความแปรปรวนของค่าความสามารถของผู้สอบ ระหว่างผู้เรียนภายในโรงเรียนเดียวกัน ระหว่างโรงเรียน และระหว่างสังกัด กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในเขตกรุงเทพมหานคร ที่ได้จากการสุ่มแบบหลายขั้นตอนจำนวน 1,002 คน จาก 24 โรงเรียน ที่มาจากสังกัด สพฐ. สช. และ กทม. เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ มีค่าความเที่ยงของแบบสอบเท่ากับ 0.90 และค่าความยากและค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.59 และ 0.58 ตามลำดับ ในการวิเคราะห์ข้อมูลใช้สถิติบรรยายและวิเคราะห์ข้อสอบ 3 แบบ ได้แก่ การวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิมด้วยโปรแกรม TAP การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบด้วยโปรแกรม BILOG MG และวิเคราะห์ข้อสอบแบบพหุระดับและวิเคราะห์ความแปรปรวนด้วยโปรแกรม HLM

#### ผลการวิจัยพบว่า

1. ผลการวิเคราะห์ข้อสอบแบบพหุระดับ ในกลุ่มผู้สอบจำนวน 200, 500 และ 1,000 คน ที่ใช้แบบสอบ 20 ข้อ มีค่าความยากและค่าความสามารถของผู้สอบที่ใกล้เคียงกัน ส่วนกลุ่มผู้สอบจำนวน 100 คน มีค่าความยากที่ต่างจากกลุ่มอื่นๆ และผู้สอบจำนวน 100, 200 และ 500 คน ที่ใช้แบบสอบ 40 ข้อ มีค่าความยากและค่าความสามารถของผู้สอบใกล้เคียงกัน ส่วนจำนวนผู้สอบ 1,000 คน มีค่าความยากและความสามารถของผู้สอบสูงกว่ากลุ่มอื่นๆ เล็กน้อย ซึ่งมีค่าใกล้เคียงกับค่าที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบมากที่สุด เมื่อพิจารณาความยาวของแบบสอบ พบว่าเมื่อแบบสอบมีความยาวเพิ่มขึ้นส่งผลให้ค่าความยากและค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์แบบพหุระดับมีค่าใกล้เคียงกับค่าที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ในทุกจำนวนผู้สอบ
2. ความยากของข้อสอบที่ได้จากทั้ง 3 แบบ ส่วนใหญ่มีความสัมพันธ์กันสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกจำนวนผู้สอบและความยาวแบบสอบ ยกเว้นจำนวนผู้สอบ 100 คน ที่ใช้แบบสอบ 20 ข้อ พบความสัมพันธ์ของความยากของข้อสอบระหว่าง 3 แบบ ไม่มีนัยสำคัญทางสถิติ ส่วนความสามารถผู้สอบระหว่าง 3 แบบ มีความสัมพันธ์ทางบวกระดับสูงอย่างมีนัยสำคัญทางสถิติ ยกเว้นจำนวนผู้สอบ 200 และ 100 คน ที่ใช้แบบสอบ 20 ข้อ ที่มีความสัมพันธ์ในระดับปานกลาง
3. ค่าความยากของข้อสอบรายข้อไม่มีความผันแปรระหว่างสังกัด ร้อยละ 80 มีความผันแปรระหว่างสังกัด ร้อยละ 20 โดยทุกข้อมีขนาดความผันแปรระหว่างสังกัดสูงกว่าระหว่างโรงเรียน ส่วนค่าความสามารถของผู้สอบมีความผันแปรระหว่างโรงเรียนและไม่มีความผันแปรระหว่างสังกัดทุกจำนวนผู้สอบ ยกเว้นกลุ่มตัวอย่าง 1,000 คน ที่พบความผันแปรระหว่างสังกัด

The purposes of this research were 1) to analyze the item difficulties and examinees' abilities resulted from Multi-level item analysis 2) to analyze the relationship between item difficulties and examinees' abilities resulted from Multi-level item analysis and Classical Test Theory and the relationships between item difficulties and examinees' abilities resulted from Multi-level item analysis and Item Response Theory 3) to analyze variation of item difficulties across schools within types of school and to study variation of examinees' abilities across students within school, across schools and across type of schools. The sample were 1,002 Bangkok students which studied in Matayomsuksa 3 in 24 public private, and community schools lower secondary school in Bangkok selecting based on multi-stage random sampling. The research instrument was Mathematical achievement test. Data analyses were descriptive statistics, 3 types of item analysis based on Classical Test Theory analyzed TAP program, item analyses based on Item Response Theory analyzed BILOG MG program, item analyses based on Multi-level item analysis and analysis of variance analyzed HLM program.

The results were as follows:

1) resulted from Multi-level item analysis for sample size 200, 500 and 1,000 using 20 items had nearly item difficulties and examinees' abilities, but sample size 100 had different difficulties from other groups and sample. The item difficulties and examinees' abilities of sample size 100, 200 and 500 using 40 items were nearly the same, but sample size 1,000 had item difficulties and examinees' abilities slightly more than other groups. Those results were nearly the same with result from IRT. When test lengths had more increased the item difficulties and examinees' abilities, the results from Multi-level item analysis were nearly the same with those result from Item Response Theory for all cases of sample size.

2) Most item difficulties resulted from 3 types had high statistical relation at .05 level of statistical significance for all sample sizes and test lengths, except 100 sample size with 20 items that had not significantly statistical relation of item difficulties among 3 types, sample sizes 200 and 100 persons with 20 items had a medium relation.

3) Item difficulties of 40 items-test did not vary across types of school (80%) and vary across types of school (20%), The variance across types of school was more than variance across school for all items. And examinees' abilities vary across schools but examinees' abilities did not vary across types of school for all sample sizes except the case sample sizes 1,000 persons that examinees' abilities vary across types of school.