

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์:
การเปรียบเทียบระหว่างไฮราซิคอลลีเนียร์โมเดล พาเชียลเครดิตโมเดลและเกรดเรสปอนส์โมเดล**

**Detecting Differential Item Functioning of a Mathematics Test Anxiety Scale: a Comparison
Between Hierarchical Linear Model Partial Credit Model and Graded Response Model**

ศิริรัตน์ สุคันธพฤษ* โขติกา ภาชีผล และศิริชัย กาญจนวาสี

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ของการศึกษาเพื่อตรวจสอบการทำหน้าที่ต่างกันแบบวัดความวิตกกังวล ในการสอบคณิตศาสตร์ โดยเปรียบเทียบระหว่างไฮราซิคอลลีเนียร์โมเดล พาเชียลเครดิตโมเดลและเกรดเรสปอนส์โมเดล กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 สายวิทย์-คณิต ปีการศึกษา 2552 จำนวน 1,715 คน จาก 29 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาพระนครศรีอยุธยาเขต 1 และเขต 2 สำนักงานเขตพื้นที่ศึกษา อ่างทองและสำนักงานเขตพื้นที่ศึกษานนทบุรี ซึ่งได้มาจากการสุ่มตัวอย่างแบบยกชั้น เครื่องมือที่ใช้ในการวิจัยคือ แบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ การวิเคราะห์ข้อมูล 3 ขั้นตอน คือ 1) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model: HLM โดยใช้โปรแกรม HLM 2) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย Partial Credit Model: PCM กับ Graded Response Model: GRM ด้วยโปรแกรม PARSCALE และ 3) เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย HLM โดยใช้โปรแกรม HLM กับ PCM กับ GRM ด้วยโปรแกรม PARSCALE ผลการวิจัย พบว่า ข้อคำถามที่ทำหน้าที่ต่างกันของข้อร่วมระหว่าง HLM, PCM และ GRM มี 6 ข้อ จาก 39 ข้อ คิดเป็นร้อยละ 15.38 ข้อคำถามที่ทำหน้าที่ต่างกันของข้อร่วมระหว่าง HLM กับ PCM มี 7 ข้อ จาก 39 ข้อ คิดเป็นร้อยละ 17.94 และข้อคำถามที่ทำหน้าที่ต่างกันของข้อร่วมระหว่าง HLM กับ GRM มี 9 ข้อ จาก 39 ข้อ คิดเป็นร้อยละ 23.07

คำสำคัญ : การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไฮราซิคอลลีเนียร์โมเดล พาเชียลเครดิตโมเดล เกรดเรสปอนส์โมเดล

Abstract

The objective of this research was to compare the Differential Item Functioning (DIF) detecting in A Mathematics Test Anxiety Scale between HLM, Partial Credit Model and Graded Response Model. The samples were 1,715 Mathayomsuksa 6 students who were studying in Math-science classes in 2009 academic year.

* ผู้ประสานงานหลัก

** งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนการศึกษาจาก ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย

Samples were drawn from 29 schools under the jurisdiction of the Pranakhon Sri Ayutthaya Educational Service Area Office 1 and 2, the Anghthong Educational Service Area Office, and the Nonthaburi Educational Service Area Office using cluster random sampling technique. Data were collected by The Mathematics Test Anxiety Scale. There were three steps of data analysis: 1) the analysis of DIF by Hierarchical Linear Model: HLM using HLM program 2) the analysis of DIF by Partial Credit Model: PCM and Graded Response Model: GRM using PARSCALE 3) the comparison of results of analysis between HLM using HLM program and PCM and GRM using PARSCALE. The major findings were: The results from DIF examination detected 6 common items from 39 items between HLM, PCM and GRM (15.38%), 7 common from 39 items between HLM and PCM (17.94%) and 8 common items from 39 items between HLM and GRM (23.07%).

Keywords : Differential Item Functioning, Hierarchical Linear Model, Partial Credit Model, Graded Response Model

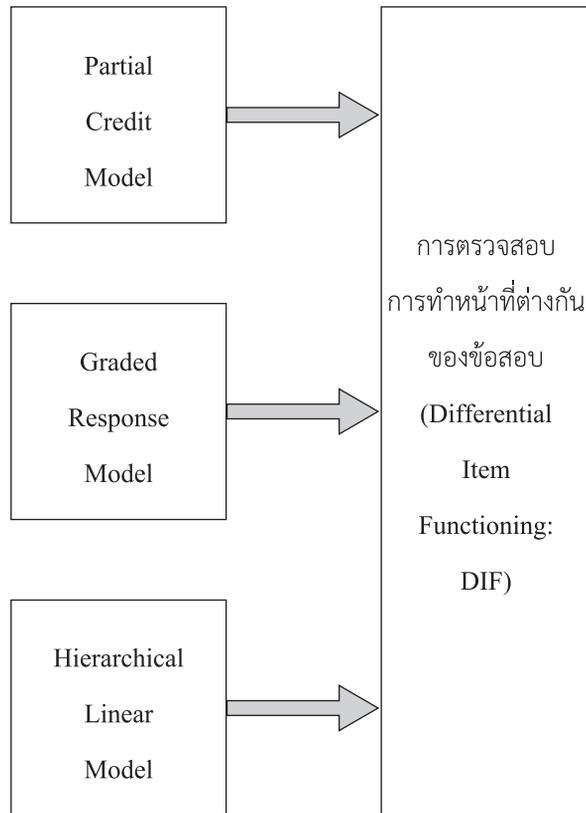
บทนำ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) คือ การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน (ศิริชัย กาญจนวาสี, 2550) ซึ่งวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีวิธีการตรวจสอบหลายวิธี ทั้งข้อสอบที่เป็นทวิภาค (Dichotomous DIF) และข้อสอบหรือข้อคำถามที่เป็นพหุภาค (Polytomous DIF) ซึ่งวิธีการตรวจสอบ DIF สำหรับข้อสอบหรือข้อคำถามที่เป็นพหุภาค เช่น วิธี General Mentel-Haenszel หรือวิธีทฤษฎีการตอบสนองข้อสอบ โดยใช้ Partial Credit Model หรือ Graded Response Model ต้องแบ่งกลุ่มเป็นสองกลุ่ม แต่การใช้ Hierarchical Linear Model ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่จำเป็นต้องแบ่งเป็นสองกลุ่ม ในกรณีที่ตัวแปรเป็นตัวแปรต่อเนื่อง ซึ่งมีความสะดวกอย่างยิ่ง ในการจัดกระทำข้อมูลก่อนตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Beretvas และ Williams, 2006) ผู้วิจัยจึงมีความสนใจที่จะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์โดยเปรียบเทียบระหว่าง Hierarchical Linear Model (HLM) โดยใช้โปรแกรม HLM กับ Partial Credit Model (PCM) และ Graded Response Model (GRM) ด้วยโปรแกรม PARSCALE โดยตรวจสอบ DIF ในส่วนของความแตกต่างระหว่างเพศ

วัตถุประสงค์

1. เพื่อตรวจสอบการทำหน้าที่ต่างกันแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM
2. เพื่อตรวจสอบการทำหน้าที่ต่างกันแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ด้วย Partial Credit Model และ Graded Response Model ด้วยโปรแกรม PARSCALE
3. เพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง Hierarchical Linear Model, Partial Credit Model และ Graded Response Model

กรอบแนวคิด



ภาพ 1 กรอบแนวคิด

วิธีการศึกษา

1. ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 แผนการเรียนวิทย์-คณิต ปีการศึกษา 2552 จังหวัดพระนครศรีอยุธยา จังหวัดอ่างทอง และจังหวัดนนทบุรี สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จำนวน 3,629 คน

2. กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 แผนการเรียนวิทย์-คณิต ปีการศึกษา 2552 จังหวัดพระนครศรีอยุธยา จังหวัดอ่างทอง และจังหวัดนนทบุรี สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จำนวน 1,715 คน ที่ได้จากการสุ่มตัวอย่างแบบยกชั้น (cluster random sampling) คือ สุ่มโรงเรียนจากแต่ละอำเภอของจังหวัดพระนครศรีอยุธยา จังหวัดอ่างทอง และจังหวัดนนทบุรี ทั้ง 29 อำเภอ ละ 1 โรงเรียน

3. เครื่องมือที่ใช้ในการวิจัย คือ แบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ จำนวน 39 ข้อ โดยวัดใน 2 องค์ประกอบ คือ ความกังวล และสภาวะทางอารมณ์ และวัดใน 4 มิติ คือ ความกังวลก่อนสอบ ความกังวลขณะสอบ สภาวะทางอารมณ์ก่อนสอบ และสภาวะทางอารมณ์ขณะสอบ

ผลการวิจัย

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM พบว่า ข้อคำถามจำนวน 17 ข้อ คือ ข้อที่ 1, 2, 3, 6, 17, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37 และ 38 ที่มีอิทธิพลของตัวแปรความเป็นเพศหญิงมีผลทำให้สัมประสิทธิ์ความชันมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งหมายความว่า ตัวแปรความเป็นเพศหญิงส่งผลให้เกิดโอกาสในการตอบข้อที่ 1, 2, 3, 6, 17, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37 และ 38 ได้แตกต่างกันระหว่างเพศในนักเรียนแต่ละคน โดยข้อที่ 17, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37 และ 38 นั้น เพศหญิงจะสามารถตอบข้อคำถามได้สูงกว่าเพศชาย ข้อคำถามที่ 1, 2, 3 และ 6 เพศชายจะสามารถตอบข้อคำถามได้สูงกว่าเพศหญิง รายละเอียดดังตาราง 1

ตาราง 1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM

Fixed effect	การประมาณค่าพารามิเตอร์ของข้อคำถาม				S.E.	coefficient	t-ratio	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4					
γ_{01}	3.532	5.383	7.655	9.832	0.157	-0.360	-2.297	66765	0.022
γ_{02}	3.530	5.382	7.653	9.831	0.156	-0.364	-2.343	66765	0.019
γ_{03}	4.415	6.266	8.538	10.716	0.158	-0.370	-2.332	66765	0.020
γ_{04}	4.449	6.301	8.572	10.750	0.156	-0.144	-0.919	66765	0.359
γ_{05}	3.977	5.828	8.100	10.278	0.158	-0.218	-1.382	66765	0.167
γ_{06}	4.989	6.841	9.112	11.290	0.157	-0.431	-2.745	66765	0.006
γ_{07}	3.111	4.963	7.234	9.412	0.154	0.095	0.620	66765	0.535
γ_{08}	4.210	6.061	8.333	10.510	0.158	-0.033	-0.209	66765	0.834
γ_{09}	4.408	6.260	8.531	10.709	0.158	-0.277	-1.748	66765	0.080
γ_{10}	3.028	4.880	7.151	9.329	0.157	0.225	1.439	66765	0.150
γ_{11}	1.787	3.638	5.910	8.087	0.160	-0.191	-1.192	66765	0.234
γ_{12}	2.943	4.794	7.066	9.243	0.157	-0.232	-1.481	66765	0.139
γ_{13}	3.456	5.308	7.579	9.757	0.155	-0.262	-1.686	66765	0.091
γ_{14}	3.239	5.091	7.362	9.54	0.152	0.024	0.159	66765	0.874
γ_{15}	3.4600	5.311	7.583	9.761	0.153	0.007	0.047	66765	0.963
γ_{16}	4.594	6.445	8.716	10.894	0.156	-0.290	-1.863	66765	0.062
γ_{17}	2.526	4.378	6.649	8.827	0.149	0.308	2.066	66765	0.038

ตาราง 1 (ต่อ)

Fixed effect	การประมาณค่าพารามิเตอร์ของข้อคำถาม				S.E.	coefficient	t-ratio	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4					
γ_{18}	3.658	5.509	7.781	9.958	0.156	0.107	0.688	66765	0.491
γ_{19}	3.788	5.640	7.911	10.089	0.150	0.002	0.016	66765	0.988
γ_{20}	3.663	5.515	7.786	9.964	0.149	-0.079	-0.528	66765	0.597
γ_{21}	2.476	4.327	6.599	8.776	0.155	0.104	0.674	66765	0.500
γ_{22}	2.766	4.618	6.889	9.067	0.159	-0.032	-0.202	66765	0.841
γ_{23}	2.424	4.275	6.547	8.725	0.148	0.050	0.338	66765	0.735
γ_{24}	2.576	4.428	6.699	8.877	0.149	0.046	0.306	66765	0.760
γ_{25}	0.617	2.469	4.740	6.918	0.161	0.659	4.079	66765	0.000
γ_{26}	0.479	2.330	4.602	6.78	0.158	0.434	2.738	66765	0.007
γ_{27}	-0.760	1.092	3.363	5.541	0.178	0.842	4.733	66765	0.000
γ_{28}	1.045	2.897	5.168	7.346	0.162	0.677	4.167	66765	0.000
γ_{29}	2.251	4.103	6.374	8.552	0.151	0.068	0.453	66765	0.650
γ_{30}	0.969	2.820	5.092	7.270	0.151	0.095	0.626	66765	0.531
γ_{31}	0.166	2.017	4.289	6.467	0.155	0.499	3.223	66765	0.002
γ_{32}	-0.079	1.773	4.044	6.222	0.147	0.490	3.344	66765	0.001
γ_{33}	0.712	2.564	4.835	7.013	0.135	0.272	2.011	66765	0.044
γ_{34}	-1.072	0.779	3.051	5.228	0.160	0.748	4.675	66765	0.000
γ_{35}	-0.888	0.964	3.235	5.413	0.153	0.714	4.664	66765	0.000
γ_{36}	-1.217	0.635	2.906	5.084	0.155	0.724	4.668	66765	0.000
γ_{37}	-0.726	1.125	3.396	5.574	0.144	0.699	4.838	66765	0.000
γ_{38}	0.357	2.209	4.480	6.658	0.140	0.441	3.142	66765	0.002
γ_{39}	1.721	3.573	5.844	8.022	0.136	-0.260	-1.911	1712	0.056

2. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย Partial Credit Model และ Graded Response Model ด้วยโปรแกรม PARSCALE

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Partial Credit Model ด้วยโปรแกรม PARSCALE พบว่า ข้อคำถามจำนวน 14 ข้อ คือ ข้อที่ 2, 6, 9, 10, 16, 17, 21, 23, 25, 26, 27, 28, 29 และ 30 ที่อิทธิพลของตัวแปรความเป็นเพศหญิงมีผลทำให้สัมประสิทธิ์ความชันมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งหมายความว่า ตัวแปรความเป็นเพศหญิงส่งผลให้เกิดโอกาสในการตอบข้อที่ 2, 6, 9, 10, 16, 17, 21, 23, 25, 26, 27, 28, 29 และ 30 ได้แตกต่างกันระหว่างเพศในนักเรียนแต่ละคน โดยข้อที่ 10, 17, 21, 23, 25, 27 และ 28 นั้นเพศหญิงจะสามารถตอบข้อคำถามได้สูงกว่าเพศชาย ข้อคำถามที่ 2, 6, 9, 16, 26, 29, และ 30 เพศชายจะสามารถตอบข้อคำถามได้สูงกว่าเพศหญิง รายละเอียดดังตาราง 2

ตาราง 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ด้วย Partial Credit Model โดยใช้โปรแกรม PARSCALE

ข้อที่	การประมาณค่าพารามิเตอร์ของข้อคำถาม				contrast	χ^2	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4				
1	1.532	0.629	-0.731	-1.429	0.049	0.397	1	0.536
2	1.604	0.576	-0.606	-1.574	-0.256	12.213	1	0.001
3	1.490	0.652	-0.560	-1.582	-0.095	1.498	1	0.219
4	1.643	0.573	-0.488	-1.728	-0.062	0.554	1	0.463
5	1.624	0.573	-0.670	-1.526	0.001	0.000	1	0.939
6	1.538	1.144	-0.475	-2.207	-0.516	13.233	1	0.000
7	2.036	0.812	-0.980	-1.869	0.155	1.448	1	0.227
8	1.679	1.022	-0.783	-1.918	0.003	0.001	1	0.930
9	1.685	0.627	-0.580	-1.733	-0.179	5.096	1	0.023
10	1.828	0.527	-0.769	-1.587	0.247	6.853	1	0.009
11	1.108	0.411	-0.497	-1.022	-0.021	0.065	1	0.787
12	1.330	0.517	-0.503	-1.343	-0.049	0.496	1	0.488
13	1.481	0.609	-0.522	-1.569	-0.021	0.097	1	0.750
14	1.595	0.527	-0.661	-1.461	0.118	2.478	1	0.111
15	1.921	0.687	-0.842	-1.766	0.065	0.576	1	0.454
16	2.401	1.379	-1.252	-2.527	-0.485	11.391	1	0.001
17	2.259	0.757	-1.051	-1.964	0.328	11.998	1	0.001
18	1.906	0.564	-0.584	-1.886	0.223	3.729	1	0.051

ตาราง 2 (ต่อ)

ข้อที่	การประมาณค่าพารามิเตอร์ของข้อคำถาม				contrast	χ^2	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4				
19	1.937	0.612	-0.635	-1.914	0.059	0.586	1	0.450
20	1.560	0.577	-0.544	-1.593	0.041	0.351	1	0.561
21	1.487	0.410	-0.560	-1.337	0.170	4.786	1	0.027
22	1.388	0.427	-0.573	-1.242	0.084	1.225	1	0.268
23	1.504	0.454	-0.645	-1.312	0.141	3.923	1	0.045
24	1.969	0.674	-0.919	-1.724	-0.198	2.857	1	0.087
25	1.195	0.465	-0.596	-1.064	0.246	5.204	1	0.021
26	1.289	0.432	-0.645	-1.076	-0.296	6.483	1	0.011
27	1.129	0.541	-0.415	-1.255	0.447	8.542	1	0.004
28	1.263	0.604	-1.154	-0.714	0.390	7.311	1	0.007
29	1.380	0.579	-0.792	-1.167	-0.310	8.197	1	0.004
30	0.970	0.548	-0.407	-1.111	-0.303	7.876	1	0.005
31	1.079	0.412	-0.420	-1.071	0.026	0.106	1	0.741
32	1.031	0.410	-0.397	-1.044	0.088	0.856	1	0.358
33	0.859	0.287	-0.292	-0.854	-0.135	1.726	1	0.185
34	1.462	0.612	-0.374	-1.700	0.023	0.113	1	0.733
35	1.354	0.599	-0.619	-1.334	0.099	2.192	1	0.134
36	1.447	0.433	-0.536	-1.355	0.049	0.416	1	0.526
37	1.237	0.408	-0.453	-1.192	0.100	1.300	1	0.253
38	0.917	0.510	-0.309	-1.118	0.007	0.004	1	0.904
39	0.820	0.428	-0.444	-0.805	-0.233	3.243	1	0.068

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Graded Response Model ด้วยโปรแกรม PARSCALE พบว่า ข้อคำถามจำนวน 19 ข้อ คือ ข้อที่ 1, 3, 6, 10, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 36 และ 39 ที่อิทธิพลของตัวแปรความเป็นเพศหญิงมีผลทำให้สัมประสิทธิ์ความชันมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งหมายความว่า ตัวแปรความเป็นเพศหญิงส่งผลให้เกิดโอกาสในการตอบข้อที่ 1, 3, 6, 10, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 36 และ 39 ได้แตกต่างกันระหว่างเพศในนักเรียนแต่ละคน โดยข้อที่ 10, 17, 19, 21, 22, 23, 25, 27 และ 28 นั้นเพศหญิงจะสามารถตอบข้อคำถามได้สูงกว่าเพศชาย ข้อคำถามที่ 1, 3, 6, 16, 24, 26, 29, 30, 36 และ 39 เพศชายจะสามารถตอบข้อคำถามได้สูงกว่าเพศหญิง รายละเอียดดังตาราง 3

ตาราง 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ด้วย Graded Response Model โดยใช้โปรแกรม PARSCALE

ข้อที่	การประมาณค่าพารามิเตอร์ของข้อคำถาม				contrast	χ^2	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4				
1	1.954	0.808	-0.827	-1.935	-0.224	8.045	1	0.005
2	2.001	0.759	-0.715	-2.045	0.021	0.079	1	0.770
3	1.936	0.804	-0.689	-2.051	-0.214	7.567	1	0.006
4	2.076	0.770	-0.603	-2.243	0.061	0.535	1	0.471
5	2.029	0.761	-0.760	-2.030	-0.099	1.688	1	0.191
6	3.040	1.363	-0.931	-3.472	-0.621	22.104	1	0.000
7	3.037	1.138	-1.143	-3.032	0.075	0.409	1	0.530
8	2.961	1.270	-1.055	-3.176	-0.242	3.033	1	0.078
9	2.112	0.795	-0.691	-2.216	-0.036	0.207	1	0.654
10	2.521	0.805	-0.906	-2.420	0.263	8.379	1	0.004
11	1.651	0.530	-0.633	-1.548	-0.043	0.292	1	0.596
12	1.670	0.645	-0.610	-1.706	0.036	0.272	1	0.609
13	1.813	0.762	-0.621	-1.953	0.038	0.319	1	0.579
14	2.060	0.687	-0.784	-1.963	0.116	2.354	1	0.121
15	2.537	0.908	-0.979	-2.467	0.082	0.939	1	0.334
16	3.693	1.563	-1.494	-3.762	-0.607	21.631	1	0.000
17	3.097	1.044	-1.204	-2.937	0.376	15.872	1	0.000
18	2.824	0.935	-0.872	-2.887	0.206	3.272	1	0.067
19	2.364	0.803	-0.755	-2.412	0.154	3.961	1	0.044
20	1.906	0.725	-0.640	-1.991	0.084	1.483	1	0.221
21	1.938	0.560	-0.693	-1.805	0.253	11.055	1	0.001
22	1.840	0.585	-0.689	-1.736	0.150	3.874	1	0.046
23	1.877	0.579	-0.734	-1.722	0.172	5.999	1	0.014
24	2.874	0.950	-1.096	-2.728	-0.332	7.173	1	0.007
25	2.070	0.638	-0.792	-1.917	0.432	18.627	1	0.000
26	1.707	0.502	-0.716	-1.493	-0.231	4.301	1	0.036
27	1.457	0.413	-0.546	-1.324	0.673	25.997	1	0.000
28	2.513	0.846	-1.157	-2.201	0.662	21.500	1	0.000

ตาราง 3 (ต่อ)

ข้อที่	การประมาณค่าพารามิเตอร์ของข้อคำถาม				contrast	χ^2	df	p-value
	Threshold1	Threshold2	Threshold3	Threshold4				
29	2.195	0.760	-0.877	-2.078	-0.268	5.405	1	0.019
30	1.627	0.656	-0.539	-1.744	-0.346	10.224	1	0.002
31	1.222	0.441	-0.456	-1.207	-0.009	0.012	1	0.877
32	1.296	0.461	-0.504	-1.253	0.107	1.167	1	0.280
33	1.240	0.414	-0.400	-1.254	-0.115	1.658	1	0.195
34	1.459	0.454	-0.394	-1.519	-0.015	0.041	1	0.821
35	1.394	0.443	-0.539	-1.298	0.000	0.000	1	0.943
36	1.135	0.403	-0.442	-1.096	-0.172	4.559	1	0.031
37	1.181	0.412	-0.388	-1.205	0.015	0.028	1	0.842
38	1.309	0.499	-0.438	-1.369	-0.059	0.336	1	0.569
39	1.444	0.468	-0.538	-1.375	-0.343	8.564	1	0.004

ผลการตรวจพบการทำหน้าที่ต่างกันของข้อสอบของทั้งสามโมเดล คือ HLM, PCM และ GRM ทั้งสิ้น 28 ข้อ รายละเอียดดังตาราง 5

ตาราง 5 ผลการเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย HLM, PCM และ GRM

ข้อที่	HLM	PCM	GRM
1. ฉันกลัวการสอบวิชาคณิตศาสตร์	✓		✓
2. ฉันรู้สึกกระวนกระวายใจ เมื่อต้องสอบวิชาคณิตศาสตร์	✓	✓	
3. ฉันคิดว่าจะได้คะแนนวิชาคณิตศาสตร์น้อย	✓		✓
6. ฉันไม่เคยกังวลในการสอบวิชาคณิตศาสตร์	✓	✓	✓
9. ฉันคิดว่าคงทำคะแนนสอบวิชาคณิตศาสตร์ ได้ไม่ดี		✓	
10. ฉันมักจะอ่านหนังสือไม่ตรงกับข้อสอบวิชาคณิตศาสตร์		✓	✓
16. ฉันสามารถแก้ปัญหาโจทย์วิชาคณิตศาสตร์ได้ดี		✓	✓
17. ฉันตอบข้อสอบวิชาคณิตศาสตร์ไม่ตรงประเด็นคำถาม	✓	✓	✓
19. ฉันเกิดความสับสนในการทำข้อสอบวิชาคณิตศาสตร์			✓
21. ฉันหงุดหงิดเมื่อจะสอบวิชาคณิตศาสตร์		✓	✓
22. ฉันรู้สึกเบื่อหน่ายเมื่อต้องสอบวิชาคณิตศาสตร์			✓
23. ฉันรู้สึกหมกมุ่นใจที่จะสอบวิชาคณิตศาสตร์		✓	✓
24. ฉันมักไม่มีสมาธิ ในวันที่ต้องทำข้อสอบวิชาคณิตศาสตร์			✓

ตาราง 5 (ต่อ)

ข้อที่	HLM	PCM	GRM
25. ฉันมักจะแยกตัว ไม่อยากพบปะผู้คน ในวันสอบวิชาคณิตศาสตร์	✓	✓	✓
26. ฉันมักจะฉุนเฉียว โโมโหง่าย ในวันสอบวิชาคณิตศาสตร์	✓	✓	✓
27. ฉันมีเรื่องทะเลาะกับเพื่อน ในวันสอบวิชาคณิตศาสตร์	✓	✓	✓
28. ฉันปล่อยปลະละเลยไม่สนใจตนเอง ในวันสอบวิชาคณิตศาสตร์	✓	✓	✓
29. ฉันหวาดกลัวการสอบวิชาคณิตศาสตร์		✓	✓
30. มือเย็น ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์		✓	✓
31. มือสั่น ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
32. ปวดท้อง ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
33. อ่อนเพลีย ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
34. เจ็บหน้าอก ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
35. หายใจติดขัด ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
36. หายใจไม่ออก ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		✓
37. มีอาการเกร็งของกล้ามเนื้อ ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
38. หัวใจเต้นเร็วผิดปกติ ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์	✓		
39. เวียนศีรษะ ในขณะที่กำลังทำข้อสอบวิชาคณิตศาสตร์		✓	

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ระหว่าง HLM กับ PCM พบว่า จำนวนข้อร่วมของการเกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง HLM กับ PCM มีจำนวน 7 ข้อ คิดเป็นร้อยละ 17.94 จากจำนวนข้อคำถาม 39 ข้อ รายละเอียดดังตาราง 6

ตาราง 6 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง Hierarchical Linear Model กับ Partial Credit Model

		HLM	
		เกิด DIF	ไม่เกิด DIF
PCM	เกิด DIF	7	7
	ไม่เกิด DIF	10	15

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ระหว่าง HLM กับ GRM พบว่า จำนวนข้อร่วมของการเกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง HLM กับ GRM มีจำนวน 9 ข้อ คิดเป็นร้อยละ 23.07 จากจำนวนข้อคำถาม 39 ข้อ รายละเอียดดังตาราง 7

ตาราง 7 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง Hierarchical Linear Model กับ Graded Response Model

		HLM	
		เกิด DIF	ไม่เกิด DIF
GRM	เกิด DIF	9	10
	ไม่เกิด DIF	8	12

ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ระหว่าง PCM กับ GRM พบว่า จำนวนข้อร่วมของการเกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง PCM กับ GRM มีจำนวน 12 ข้อ คิดเป็นร้อยละ 30.76 จากจำนวนข้อคำถาม 39 ข้อ รายละเอียดดังตาราง 8

ตาราง 8 จำนวนข้อของการเกิดและไม่เกิดการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ระหว่าง Partial Credit Model กับ Graded Response Model

		GRM	
		เกิด DIF	ไม่เกิด DIF
PCM	เกิด DIF	12	2
	ไม่เกิด DIF	7	18

สรุปผลการศึกษา

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM พบว่า ข้อคำถามจำนวน 17 ข้อ คือ ข้อที่ 1, 2, 3, 6, 17, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37 และ 38 เกิดการทำหน้าที่ต่างกันของข้อสอบ

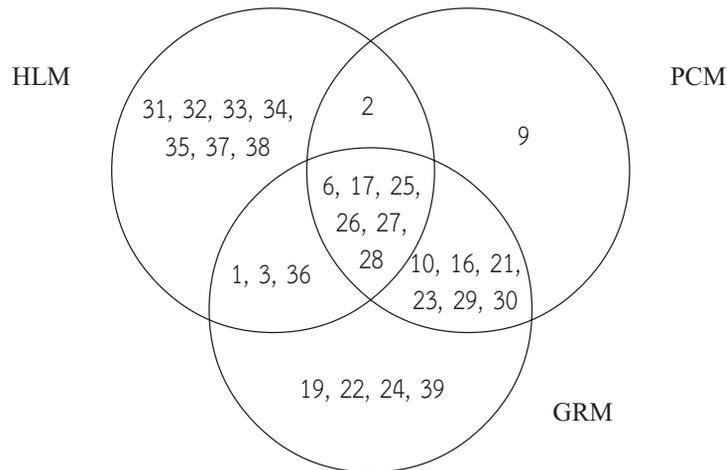
2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Partial Credit Model ด้วยโปรแกรม PARSCALE พบว่า ข้อคำถามจำนวน 14 ข้อ คือ ข้อที่ 2, 6, 9, 10, 16, 17, 21, 23, 25, 26, 27, 28, 29 และ 30 เกิดการทำหน้าที่ต่างกันของข้อสอบ

3. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Graded Response Model ด้วยโปรแกรม PARSCALE พบว่า ข้อคำถามจำนวน 19 ข้อ คือ ข้อที่ 1, 3, 6, 10, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 36 และ 39 เกิดการทำหน้าที่ต่างกันของข้อสอบ

4. ผลการเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่าง HLM, PCM และ GRM พบว่า ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบร่วมระหว่างด้วย HLM โดยใช้โปรแกรม HLM กับ PCM และ Graded Response Model โดยใช้โปรแกรม PARSCALE มี 6 ข้อ คือ ข้อที่ 6, 17, 25, 26, 27 และ 28 ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบร่วมระหว่าง Hierarchical Linear Model โดยใช้โปรแกรม HLM กับ PCM โดยใช้โปรแกรม PARSCALE มี 1 ข้อ คือข้อ 2 ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบร่วมระหว่าง HLM โดยใช้โปรแกรม HLM กับ GRM โดยใช้โปรแกรม PARSCALE มี 3 ข้อ คือ ข้อที่ 1, 3 และ 36 และข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบร่วมระหว่าง PCM กับ GRM โดยใช้โปรแกรมใช้ PARSCALE ด้วย มี 6 ข้อ คือ ข้อที่ 10, 16, 21, 23, 29 และ 30 ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM มี 7 ข้อ คือข้อที่ 31, 32, 33, 34, 35, 37 และ 38 ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบด้วย PCM โดยใช้โปรแกรม PARSCALE มี 1 ข้อ คือข้อที่ 9 และ ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบด้วย Graded Response Model โดยใช้โปรแกรม PARSCALE มี 4 ข้อ คือข้อที่ 19, 22, 24 และ 39 รายละเอียดดังตาราง 4 และภาพ 2

ตาราง 4 ผลการเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย Hierarchical Linear Model โดยใช้โปรแกรม HLM กับ Partial Credit Model และ Graded Response Model โดยใช้โปรแกรม PARSCALE

โปรแกรม	ข้อคำถามที่ทำหน้าที่ต่างกันของข้อสอบ	ข้อร่วมการเกิด DIF ระหว่าง HLM, PCM และ GRM	ข้อร่วมการเกิด DIF ระหว่าง HLM กับ PCM	ข้อร่วมการเกิด DIF ระหว่าง HLM กับ GRM	ข้อร่วมการเกิด DIF ระหว่าง PCM กับ GRM
HLM	1, 2, 3, 6, 17, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37 และ 38				
PCM	2, 6, 9, 10, 16, 17, 21, 23, 25, 26, 27, 28, 29 และ 30	6, 17, 25, 26, 27 และ 28	2	1, 3 และ 36	10, 16, 21, 23, 29 และ 30
GRM	1, 3, 6, 10, 16, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 36 และ 39				



ภาพ 2 ข้อที่เกิด DIF ร่วมกันระหว่าง HLM, PCM และ GRM

อภิปรายผลการศึกษา

1. ในเรื่องของเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ HLM, PCM และ GRM พบว่าให้ผลการตรวจสอบการทำหน้าที่ของข้อสอบแตกต่างกัน จึงยังสรุปไม่ได้ว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีใดดีกว่าวิธีใด ซึ่งหากต้องการทราบว่าวิธีใดดีกว่า ต้องทำในสถานการณ์ที่มีการจำลองข้อมูลขึ้นมาและตั้งเกณฑ์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2. จากผลการวิจัย พบว่า HLM โดยใช้โปรแกรม HLM กับ PCM และ GRM โดยใช้โปรแกรม PARSCALE สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความวิตกกังวลในการสอบคณิตศาสตร์ได้ทั้งสามโมเดล โดยข้อคำถามที่ทำหน้าที่ต่างกันของร่วมระหว่าง HLM โดยใช้โปรแกรม HLM กับ PCM และ GRM โดยใช้โปรแกรม PARSCALE มี 6 ข้อ คือ ข้อที่ 6, 17, 25, 26, 27 และ 28 เมื่อพิจารณาเนื้อความใน ข้อคำถาม มีดังนี้

ข้อที่ 6 ฉันไม่เคยกังวลในการสอบคณิตศาสตร์

ข้อที่ 17 ฉันตอบข้อสอบวิชาคณิตศาสตร์ไม่ตรงประเด็น

ข้อที่ 25 ฉันมักจะแยกตัว ไม่อยากพบปะผู้คน ในวันสอบวิชาคณิตศาสตร์

ข้อที่ 27 ฉันมีเรื่องทะเลาะกับเพื่อน ในวันสอบวิชาคณิตศาสตร์

ข้อที่ 28 ฉันปล่อยปละละเลยไม่สนใจตนเอง ในวันสอบวิชาคณิตศาสตร์

จากข้อคำถามทั้ง 6 ข้อ พบว่า ข้อคำถามที่ 6 นักเรียนชายจะตอบข้อคำถามได้สูงกว่านักเรียนหญิง ส่วนข้อ 17, 25, 27 และ 28 นักเรียนหญิงจะตอบข้อคำถามได้สูงกว่านักเรียนชาย ทั้งนี้อาจเนื่องจาก ข้อคำถามที่ 6 เป็นข้อคำถามนิเสธ นักเรียนชายอาจใช้ความคิดในการตอบมากกว่านักเรียนหญิง มีการพิจารณาข้อคำถามอย่างละเอียด จึงทำให้ตอบข้อคำถามข้อนี้ได้สูงกว่านักเรียนหญิง ส่วนข้อ 17, 25, 27 และ 28 เป็นข้อคำถามที่มีลักษณะอ่อนไหวทางอารมณ์ นักเรียนหญิงซึ่งมีลักษณะอารมณ์ที่อ่อนไหวกว่านักเรียนชาย ทำให้ใช้ความคิดในการตัดสินใจมากกว่า จึงตอบข้อคำถามเหล่านี้ได้สูงกว่า ส่วนข้อคำถามในข้อที่ 26 ไม่ได้นำมาพิจารณา เนื่องจาก ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย HLM กับ GRM ให้ผลตรงข้ามกับ PCM

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า ทั้งสามโมเดลสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ จึงสามารถนำโมเดลการตรวจสอบการทำหน้าที่ต่างกันไปใช้ได้ทั้งสามโมเดล สำหรับโมเดลที่ไม่ยุ่งยากในการวิเคราะห์ จะเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย HLM โดยใช้โปรแกรม HLM เนื่องจากสามารถศึกษาได้จากคู่มือ การอ่านค่าต่างๆ ที่ได้จากการวิเคราะห์ไม่ยุ่งยาก หรือซับซ้อน ตลอดจนไม่มีความจำเป็นต้องเขียนคำสั่งเข้าไปในโปรแกรม เมื่อการวิเคราะห์ด้วยแม่ PCM และ GRM โดยใช้โปรแกรม PARSCALE นอกจากนี้ การอ่านผลการวิเคราะห์จากโปรแกรม PARSCALE มีความยุ่งยาก ทั้งยังต้องมีการคำนวณค่าต่างๆ อีกด้วย

ข้อเสนอในการวิจัยครั้งต่อไป

1. จากการวิจัย พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งสามโมเดล คือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย HLM โดยใช้โปรแกรม HLM กับ PCM และ GRM โดยใช้โปรแกรม PARSCALE สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ แต่ได้ผลการตรวจสอบต่างกัน ดังนั้นในการวิจัยครั้งต่อไป ควรมีการจำลองข้อมูลขึ้นมาแล้วกำหนดให้มีข้อที่มีการทำหน้าที่ต่างกันของข้อสอบ แล้วใช้โมเดลการตรวจสอบทั้งสามโมเดลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แล้วดูว่าโมเดลใดสามารถตรวจสอบได้ถูกต้อง ตรงกับข้อมูลที่มีการจำลองขึ้นมา เช่น จำลองข้อมูล 20 ข้อ กำหนดให้เกิดการทำหน้าที่ต่างกันของข้อสอบ 6 ข้อ โดยกำหนดขนาดของ DIF แตกต่างกันในระดับน้อย ปานกลาง และมากอย่างละ 2 ข้อ แล้วใช้ทั้งสามโมเดลตรวจสอบ ดูว่าโมเดลใดสามารถตรวจสอบ DIF ได้ถูกต้อง ทั้งข้อที่เกิด DIF และขนาดของการเกิด DIF และความไวในการตรวจสอบ DIF

2. จากการวิจัยพบว่า จำนวนข้อคำถามที่เกิดการทำหน้าที่ต่างกันของข้อสอบมีจำนวน 28 ข้อ จากจำนวนข้อคำถามทั้งสิ้น 39 ข้อ จะเห็นได้ว่า ข้อคำถามจำนวนมากที่สร้างขึ้นทำให้เกิด DIF ดังนั้นในการศึกษา ครั้งต่อไป ควรระมัดระวังในการสร้างข้อคำถามเพื่อไม่ให้เกิดการทำหน้าที่ต่างกันของข้อสอบในแต่ละข้อคำถาม โดยการศึกษา โครงสร้าง ทฤษฎีของเรื่องที่จะทำการศึกษาอย่างลึกซึ้ง สร้างข้อคำถามที่คำนึงถึงความแตกต่างระหว่างบุคคล เพื่อที่จะไม่ให้เกิดการทำหน้าที่ต่างกันของข้อสอบ

เอกสารอ้างอิง

- กฤษณา ศักดิ์ศรี. (2530). *จิตวิทยาการศึกษา*. กรุงเทพฯ : บารุงสาสน์การพิมพ์.
- รังรอง งามศิริ. (2540) *การวิเคราะห์ความสัมพันธ์เชิงสาเหตุของตัวแปรที่มีผลต่อความวิตกกังวลในการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 กรุงเทพมหานคร*. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต ภาควิชาจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ล้วน สายยศ และ อังคณา สายยศ. (2543). *การวัดด้านจิตพิสัย*. กรุงเทพฯ : สุวีริยาสาส์น.
- ศิริชัย กาญจนวาสี, ทวีวัฒน์ ปิตยานนท์ และ ดิเรก ศรีสุขโข (2544). *การเลือกใช้สถิติที่เหมาะสมสำหรับการวิจัย*. กทม. : บริษัท บุญศิริการพิมพ์ จำกัด.

- ศิริชัย กาญจนวาสี. (2548). การวิเคราะห์พระดัต. กทม. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2548). ทฤษฎีการทดสอบแบบดั้งเดิม. กทม. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2549). การวิเคราะห์ข้อสอบแบบพระดัต. *วารสารครุศาสตร์*. 3. 42-52.
- ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่. กทม. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อานันต์ชนก วิจิตรนิเทศ. (2546). การวิเคราะห์ความสัมพันธ์เชิงสาเหตุของตัวแปรที่มีอิทธิพลต่อความวิตกกังวลในการสอบวิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- เอมอร จังศิริพรกรณ์. (2548). การเปรียบเทียบคุณภาพของแบบสอบระหว่างแบบสอบเลือกตอบที่มีจำนวนตัวถูกตัวเดียวกับตัวถูกมากกว่า 1 ตัว เมื่อตรวจด้วยวิธีการให้คะแนนความรู้อ่างส่วน. รายงานวิจัย.
- Beretvas, S. N., & Williams, N.J. (2006). DIF Identification Using HGLM for Polytomous Items. *Applied Psychological Measurement*. 30. 22-42.
- Cheong Yuk Fai. (2006). Analysis of School Context Effects on Differential Item Functioning Using Hierarchical Generalized Linear Models. *International Journal of Testing*. 1. 57-79.
- Embretson, E. S., & Reise, P. S. (2000). *Item Response Theory For Psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hambleton, R. K., & Swaminathan, A. (1985). *Item Response Theory*. MA: Kluwer Academic Publishers.
- Holland, P.W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Kamata, A. (1998, April). **One-parameter Hierarchical Generalized Linear Logistic Model: An Application of HGLM to IRT**. Paper Presented at the annual meeting of American Educational Research Association, San Diego, CA.
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*. 38. 79-93.
- Kamata, A., Chaimongkol, S., Genc, E., & Bilir, K. (2005, April). **Random-Effect Differential Item Functioning Across Group Unites by the Hierarchical Generalized Linear Model**. Paper Presented at the annual meeting of American Educational Research Association, Montreal, Canada.
- Kim, H. S., Cohen, S. A., & Kim, S. (2007). DIF Detection and Effect Size Measures for Polytomously Scored Items. *Journal of Educational Measurement*. 44. 93-116.
- Luppescu, S. (2002, April). **DIF Detection in HLM**. Paper Presented at the annual meeting of American Educational Research Association, New Orleans, LA.
- Maier, K. S. (2001). A Rasch Hierarchical Measurement Model. *Journal of Educational and Behavioral Statistic*. 26- 307-330.

- Pastor, D. A. (2003). The Use of Multilevel Item Response Theory Modeling in Applied Research: An Illustration. **Applied Measurement in Education**. 16. 223-243.
- Raudenbush, S. W., & Bryk, A. S. (2002). **Hierarchical Linear Models: Applications and Data Analysis Method**. London: SAGE Publications.
- Sawaminathan, H., & Rogers, H.J. (1990). Detecting defferential item functioning using logistic regression procedures. **Journal of Educational Measurement**. 27(4). 361-370.
- Shen, L. (1999, April). **A Multilevel Assessment of Differential Item Functioning**. Paper Presented at the annual meeting of American Educational Research Association, Montreal, Canada.
- Swanson, D. B., Clauser, B. E., Case, S. M., Nungester, R.J., & Featherman, C. (2002). Analysis of Differential Item Functioning (DIF) Using Hierarchical Logistic Regression Models. **Journal of Educational and Behavioral Statistic**, 27. 53-75.
- Vaughn, B K. (2006). **A Hierarchical Generalized Linear Model of random Differential Item Functioning for Polytomous Items: A Bayesian Multilevel Approach**. Unpublished doctoral dissertation, Florida State University, Department of Educational Psychology and Learning Systems.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of Differential Functioning for Performance Tasks. **Journal of Educational Measurement**. 30. 233-251.

คณะผู้เขียน

ศิริรัตน์ สุคันธพฤกษ์

นิสิตปริญญาเอกสาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

e-mail: siriruts@hotmail.com

รองศาสตราจารย์ ดร.โชติกา ภาษีผล

รองศาสตราจารย์ประจำคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

e-mail: aimornj@hotmail.com

ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

ศาสตราจารย์ประจำคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

e-mail: skanjanawasee@hotmail.com