

บทวิจัย

การวิเคราะห์ความลำเอียงของข้อทดสอบ ภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533

สุพัฒน์ สุขมลสันต์

ความเป็นมาและความสำคัญของปัญหา

ในแต่ละปีการศึกษา มีผู้สมัครสอบเข้าศึกษาต่อในมหาวิทยาลัยของรัฐและเอกชนจำนวนประมาณ 30,000 คน แต่จะมีจำนวนเพียงประมาณร้อยละ 25 เท่านั้นที่สามารถสอบเข้าศึกษาในมหาวิทยาลัยของรัฐได้ การที่ผู้สมัครสอบไม่สามารถสอบเข้าศึกษาต่อได้นั้น ส่วนหนึ่งเป็นเพราะมีความรู้หรือความสามารถไม่เพียงพอ เมื่อเทียบกับผู้อื่นที่สามารถสอบเข้าศึกษาต่อได้ แต่อีกส่วนหนึ่งที่สำคัญมากเกิดจากความบกพร่องของคุณภาพของข้อทดสอบที่นำมาใช้ในการทดสอบ โดยเฉพาะอย่างยิ่งวิชาภาษาอังกฤษ เพราะจากการวิเคราะห์ข้อทดสอบวิชาภาษาอังกฤษ ชุด กข. และ กขค. ของผู้รู้ทางการวัดและประเมินผล รวมทั้งทางด้านหลักสูตรในหลายปีที่ผ่านมา พบว่าแบบทดสอบดังกล่าวยังมีข้อบกพร่องมาก เช่น ในปี พ.ศ. 2533 Alec Bradford (Bradford, 1933: 17) พบว่าแบบทดสอบภาษาอังกฤษชุด กข. มีปัญหาเกี่ยวกับคำตอบ 20 ข้อหรือร้อยละ 20 เช่น ไม่มีคำตอบที่ถูกต้องและมีคำตอบมากกว่าตัวเลือก และแบบทดสอบภาษาอังกฤษ ชุด กขค. ก็มีปัญหา

เกี่ยวกับคำตอบ 18 ข้อ หรือร้อยละ 18 ในปี พ.ศ. 2527 สุพัฒน์ สุขมลสันต์ (2527: 101-104) พบว่าแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องร้อยละ 4 และ 18 และในปี 2528 สุพัฒน์ สุขมลสันต์ (2528: 1-20) ก็ยังพบว่าแบบทดสอบภาษาอังกฤษ ชุด กข. และ กขค. มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องร้อยละ 15 และ 20 ตามลำดับอีกด้วย

ดังนั้น จะเห็นได้ว่าในแต่ละปีแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยมักมีปัญหาเกี่ยวกับคำตอบที่ถูกต้องประมาณชุดละ 20 ข้อ หรือ ร้อยละ 20 ทั้งนี้ยังไม่ได้รวมกับข้อบกพร่องด้านอื่นอีกที่มีผู้ศึกษาไว้เช่นเดียวกัน เช่นความไม่สอดคล้องของหลักสูตรระดับมัธยมศึกษา กับเนื้อหาของแบบทดสอบเข้ามหาวิทยาลัย และความยากของข้อทดสอบ เป็นต้น ซึ่งข้อบกพร่องต่าง ๆ เหล่านี้ย่อมมีผลโดยตรงต่อผลการสอบเข้ามหาวิทยาลัยของผู้ที่สอบได้และสอบตก กล่าวโดยสรุปก็คือทำให้ผลการตัดสินใจ-ตก คลาดเคลื่อนได้ ความบกพร่องต่าง ๆ ดังกล่าวแล้ว ทำให้ผู้วิจัยสงสัยว่าข้อทดสอบอาจมีความบกพร่องด้านอื่นอีกด้วย เช่นความ

ลำเอียงของข้อทดสอบ (item bias) จึงมีความสนใจทำการศึกษาเรื่องนี้เพื่อเสนอผลให้ผู้ที่เกี่ยวข้องทราบ สำหรับใช้เป็นแนวทางแก้ปัญหาต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์หาความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533 โดย
 1. วิธีกำหนดจุดเดลต้า (Delta Plot Method)
 2. วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ (Chi-square Method)
 3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method)
2. เพื่อเปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ เมื่อวิเคราะห์ด้วย 3 วิธีดังกล่าว

สมมติฐานของการวิจัย

1. แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อเพศของผู้สอบ
2. แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย กข. และ กขค. ปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ
3. วิธีการทดสอบหาความลำเอียงของข้อทดสอบ 3 วิธี คือวิธีกำหนดจุดเดลต้า วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ และวิธี

การตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ น่าจะทำให้ได้จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ของผู้สอบแตกต่างกัน

ขอบเขตการวิจัย

1. วิเคราะห์หาความลำเอียงของข้อทดสอบของแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 รวม 6 ฉบับ และมีข้อสอบทั้งหมด 600 ข้อ
2. วิเคราะห์หาความลำเอียงของข้อสอบเพียง 3 วิธีดังกล่าวแล้วเท่านั้น
3. ใช้ผลการสอบของผู้สอบจริง ทั้งผู้ที่สอบเข้าศึกษาต่อในมหาวิทยาลัยได้ และเข้าศึกษาต่อไม่ได้

ความจำกัดของการวิจัย

1. คำตอบของข้อสอบบางข้อมีปัญหาเช่น ไม่มีคำตอบที่ถูกต้อง หรือมีคำตอบที่ถูกต้องมากกว่า 1 คำตอบ
2. ผู้วิจัยไม่มีค่าเฉลยอย่างเป็นทางการของข้อทดสอบแต่ละข้อ
3. แนวคิดและความหมายของคำว่า 'ความลำเอียงของข้อทดสอบ' ของการวิเคราะห์หาความลำเอียงแต่ละวิธีแตกต่างกัน
4. ศึกษาความลำเอียงต่อภาคภูมิศาสตร์โดยการเปรียบเทียบผลการสอบระหว่างผู้สอบจากภาคกลางวันภาคอื่น ๆ อีก 4 ภาคเท่านั้น (เหนือ ตะวันออก ตะวันออกเฉียงเหนือ ใต้) ไม่ได้เปรียบเทียบระหว่างภาคอื่นอีก
5. ไม่ได้มุ่งศึกษาถึงสาเหตุหรือแหล่งที่เกิด

ของความลำเอียงของข้อทดสอบแต่ละข้อ เพราะข้อทดสอบมีจำนวนมากและต้องศึกษาในหลายลักษณะ (เช่น คำนวณแล้วมีอย่างน้อยประมาณ $5 \times 100 \times 6 = 3,000$ กราฟ!)

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นข้อมูลให้แก่คณะกรรมการหรือผู้ที่รับผิดชอบในการสร้างแบบทดสอบคัดเลือกเข้ามหาวิทยาลัยปีต่อ ๆ ไป ได้ตระหนักถึงข้อบกพร่องของข้อทดสอบด้านความลำเอียงต่อเพศ และภาคภูมิศาสตร์ของผู้สอบ เพื่อหาทางปรับปรุงแก้ไข หรือป้องกันไม่ให้เกิดข้อบกพร่องดังกล่าว เพื่อให้เกิดความเป็นธรรมแก่ผู้สอบและเป็นผลดีต่อระบบการศึกษาในส่วนรวมด้วย
2. เพื่อเป็นการตรวจสอบคุณภาพของแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยในระยะ 3 ปีที่แล้ว จำนวน 6 ฉบับ อีกวิธีหนึ่งว่าแบบทดสอบดังกล่าวมีคุณภาพเป็นอย่างไร
3. เพื่อเป็นการแนะนำการวิเคราะห์หาคุณภาพของข้อทดสอบอีกวิธีหนึ่งให้นักทดสอบและผู้ที่เกี่ยวข้องหรือสนใจทั่วไปให้เห็นทั้งความสำคัญ แนวคิด และวิธีการจัดกระทำ โดยมุ่งหวังว่าจะเป็นวิธีแพร่หลายต่อไป เพื่อให้เกิดความเป็นธรรมแก่ผู้สอบและเป็นผลดีต่อระบบการศึกษาในส่วนรวมด้วย

วิธีดำเนินการ

1. ประชากร

ประชากรการวิจัยครั้งนี้ ได้แก่ผู้สอบแบบทดสอบภาษาอังกฤษเข้ามหา

วิทยาลัย ชุด กข. และ/หรือ กขค. ปี พ.ศ. 2531-2533 จำนวน 6 กลุ่ม ๆ ละประมาณ 30,000 - 80,000 คน แต่ละกลุ่มแบ่งตามเพศของผู้สอบได้ 2 กลุ่มคือ เพศชาย และเพศหญิง ซึ่งมีประมาณกลุ่มละ 8,000 - 40,000 คน และแต่ละกลุ่มแบ่งตามภาคภูมิศาสตร์ของผู้สอบได้อีก 5 กลุ่ม ซึ่งมีประมาณกลุ่มละ 1,000 - 47,000 คน ตามจำนวนผู้ที่เข้าสอบจริง

2. พลวิสัย

พลวิสัยได้แก่ตัวอย่างของประชากรของแต่ละกลุ่ม ซึ่งได้มาจากการสุ่มอย่างง่ายโดยอาศัยเครื่องคอมพิวเตอร์ช่วยในการสุ่มตามจำนวนสูงสุดเท่าที่ โปรแกรมคอมพิวเตอร์สำหรับวิเคราะห์ข้อทดสอบจะสามารถทำงานได้ (ไม่เกินกลุ่มละ 3,000 คน) และแต่ละกลุ่มมีจำนวนผู้สอบเพศชายและเพศหญิงเท่ากัน ทำให้พลวิสัยแต่ละกลุ่มมีขนาดระหว่าง 424-3,000 คน ซึ่งมีขนาดมากกว่าขนาดกลุ่มตัวอย่างที่เหมาะสมเพื่อการวิจัย เมื่อกำหนดให้มีความเชื่อมั่นในการสุ่มอย่างง่ายเท่ากับร้อยละ 99 และความคลาดเคลื่อนในการสุ่มไม่เกินร้อยละ 5 จึงเชื่อได้ว่าเป็นพลวิสัยที่มีขนาดเพียงพอ และเป็นตัวอย่างที่ดีของประชากร

3. เครื่องมือที่ใช้ในการวิจัย

ได้แก่ ผลการสอบรายข้อของข้อทดสอบจำนวน 600 ข้อ จากแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 ข้อทดสอบดังกล่าวสร้างขึ้น โดยคณะ

กรรมการสร้างแบบทดสอบของทบวงมหาวิทยาลัย ข้อทดสอบเหล่านี้เป็นแบบปรนัย ชนิดเลือกตอบ และมี 4 ตัวเลือก

4. การเก็บรวบรวมข้อมูล

1. เขียนจดหมายติดต่ออย่างเป็นทางการเพื่อขอยืมแถบแม่เหล็กบันทึกข้อมูลคอมพิวเตอร์ ที่บันทึกผลการสอบรายข้อของวิชา 05 (ภาษาอังกฤษชุด กข.) และ 10 (ภาษาอังกฤษชุด กขค.) ประจำปี 2531-2533 จากทบวงมหาวิทยาลัย

2. ดำเนินการสุ่มเลือกพลวิจัยตามขนาดและเงื่อนไขที่ต้องการโดยใช้โปรแกรมคอมพิวเตอร์

5. การวิเคราะห์ข้อมูล

1. ใช้โปรแกรมคอมพิวเตอร์ชื่อ SPSSX (The Statistical Packages for the Social Sciences, Version X) ของสถาบันบริการคอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัย เช่น ใช้ sub-program ชื่อ Sample เพื่อสุ่มตัวอย่างของประชากร ใช้ FACTOR เพื่อวิเคราะห์หาความมีมิติเดียวของข้อทดสอบ และใช้ CROSSTABS เพื่อวิเคราะห์หาความลำเอียง โดยวิธีใช้ Chi-square เป็นต้น

2. ใช้โปรแกรมคอมพิวเตอร์ชื่อ ITEMX เพื่อวิเคราะห์ข้อทดสอบรายข้อตามวิธีประเพณีนิยม และใช้โปรแกรมชื่อ LOGIST เพื่อวิเคราะห์ข้อทดสอบรายข้อตามทฤษฎีการตอบสนองของข้อทดสอบ (Item Response

Theory)

3. ใช้โปรแกรมที่ผู้วิจัยเขียนขึ้นคือ BIAS เพื่อวิเคราะห์หาความลำเอียงของข้อทดสอบตามวิธีที่ 1 และ ICCTEST เพื่อวิเคราะห์หาความลำเอียงของข้อทดสอบตามวิธีที่ 3

6. สถิติที่ใช้

ใช้สถิติที่เกี่ยวข้องหลายอย่างเช่น สถิติบรรยายการอย่างง่าย Factor Analysis, Chi-square tests, Classical Item Analysis, Logistic Item Analysis, ICC Test, Bias Test, F-test และ Simple Correlation เป็นต้น

ผลการวิจัย

1. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อเพศ โดยชุด กข. มีข้อทดสอบที่มีความลำเอียงต่อเพศโดยเฉลี่ยประมาณ 7-28 ข้อ (ร้อยละ 7-28) แล้วแต่ผลการวิเคราะห์แต่ละวิธี และชุดกขค. มีข้อทดสอบที่ลำเอียงต่อเพศโดยเฉลี่ยข้อ (ร้อยละ 4-41) ประมาณ 4-41
2. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ โดยชุด กข. มีข้อทดสอบที่มีความลำเอียงต่อภาคโดยเฉลี่ยประมาณ 6-45 ข้อ (ร้อยละ 6-45) แล้วแต่ผลการวิเคราะห์แต่ละวิธี และชุด กขค. มีข้อทดสอบที่ลำเอียงต่อภาคโดยเฉลี่ยประมาณ 5-43 ข้อ (ร้อยละ 5-43)

3. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อผู้สอบจากภาคอื่นมากกว่าจากภาคกลางประมาณ 3-4 เท่า และมีแนวโน้มว่ามีความลำเอียงต่อผู้สอบเพศชายมากกว่าเพศหญิง
4. การวิเคราะห์หาความลำเอียงของข้อทดสอบด้วย 3 วิธีที่แตกต่างกัน พบข้อทดสอบที่มีความลำเอียงจำนวนแตกต่างกันอย่างมีนัยสำคัญ และสัมพันธ์กันอย่างไม่มีนัยสำคัญ การวิเคราะห์หาเฉพาะข้อทดสอบที่ลำเอียง

ในระดับปานกลางและมากกว่านั้น พบข้อทดสอบที่ลำเอียงจำนวนมากที่สุด รองลงมา ได้แก่ วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วย ไคสแควร์ และวิธีกำหนดจุดเดลต้า พบข้อทดสอบที่ลำเอียงจำนวนน้อยที่สุด และน้อยกว่าจำนวนของวิธีที่ 1-2 ดังกล่าวแล้วประมาณ 3-4 เท่า รายละเอียดของข้อสรุปดังกล่าว พร้อมรายละเอียดบางอย่างอาจศึกษาได้จากตารางที่ 1-2 ต่อไปนี้

ตารางที่ 1 : เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของผู้สอบ เมื่อวิเคราะห์ด้วย 3 วิธี

ปี พ.ศ.	แบบทดสอบ ชุด	n ที่ลำเอียงต่อเพศ			ข้อที่ลำเอียงต่อเพศเกิน 1 วิธี	
		DPM	CSM	TPLM	n	ข้อที่
2531	กข	3	16	23	3	7, 13, 92
2532	กข	13	28	54	22	16, 19, 24, 28, 34, 39, 40, 43, 49, 51, 55-57, 63, 71, 76, 77, 81, 87, 96-98
2533	กข	4	9	8	3	59, 69, 100
\bar{X}		6.67	17.67	28.33	9.33	14.86%
2531	กขค	6	22	8	7	16, 25, 31, 76, 90, 91, 96
2532	กขค	4	17	61	15	4, 7, 8, 10, 11, 13, 19, 28, 30, 34, 55, 65, 67, 78, 84
2533	กขค	2	14	55	9	7, 12, 16, 20, 23, 30, 48, 98
\bar{X}		4.00	17.67	41.33	10.33	16.80%
\bar{X}		5.33	17.67	34.83	9.83	15.83 %
F-test		**	**	**		
r_{xy}		-	0.84*	0.18*		
			-	0.29*		

* $p > 0.05$

** $p < 0.05$

ตารางที่ 2 : เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ
เมื่อวิเคราะห์ด้วย 3 วิธี

ปี พ.ศ.	แบบทดสอบชุด	ระหว่าง ภาค	n ที่ลำเอียงต่อแต่ละภาค			ข้อที่ลำเอียงต่อภาคเกิน 1 วิธี	
			DPM	CSM	TPLM	n	ข้อที่
2531	กข	C vs E	(4) 6	(37) 40	(36) 28	14	4, 14, 16, 17, 22, 26, 33, 36, 38, 48, 50, 80, 88, 89
		C vs NE	9	41	32		
		C vs N	7	10	24		
		C vs S	8	39	34		
2532	กข	C vs E	(11) 11	(52) 55	(39) 29	28	5, 6, 10, 16-18, 25, 28, 30, 37, 38, 40, 41, 43, 45, 51, 57, 60, 64, 66, 67, 76, 81, 83, 95, 96, 98, 99
		C vs NE	10	55	53		
		C vs N	13	11	25		
		C vs S	13	48	28		
2533	กข	C vs E	(3) 5	(46) 42	(46) 45	18	10, 12-14, 19, 20, 22, 26, 30, 32, 33, 42, 64, 66, 67, 73, 91, 93
		C vs NE	6	45	55		
		C vs N	8	23	38		
		C vs S	6	36	33		
— X			(6) 8.50	(45) 37.08	(40) 35.33	20.00	21.53%

(ตารางที่ 2) : ต่อ

ปี พ.ศ.	แบบทดสอบชุด	ระหว่าง ภาค	n ที่ลำเอียงต่อแต่ละภาค			ข้อที่ลำเอียงต่อภาคเกิน 1 วิธี	
			DPM	CSM	TPLM	n	ข้อที่
2531	กขค	C vs E	(5) 8	(26) 13	(51) 60	14	2, 14, 20, 44, 48, 49, 58, 59, 62, 72, 76, 89, 90, 100
		C vs NE	4	21	42		
		C vs N	9	24	31		
		C vs S	2	32	38		
2532	กขค	C vs E	(3) 4	(36) 15	(41) 45	15	1, 12, 18-20, 24, 27, 37, 38, 57, 79, 84, 87, 91, 96
		C vs NE	4	37	43		
		C vs N	6	28	30		
		C vs S	2	33	29		
2533	กขค	C vs E	(8) 8	(18) 15	(36) 42	8	8, 15, 21, 28, 50, 56, 71, 74
		C vs NE	22	21	36		
		C vs N	9	20	19		
		C vs S	9	19	28		
\bar{X}			(5) 7.16	(27) 23.17	(43) 36.92	12.33	16.24%
\bar{X}			7.83	30.13	36.13	16.17	18.89%
F-test			**	**	**		
r_{xy}			-	0.036*	0.18*		
			-		0.13*		

* $p > 0.05$ ** $p < 0.05$

อภิปรายผล

ผลจากการวิจัยครั้งนี้มีประเด็นที่น่าสนใจหลายอย่างดังนี้

1. เหตุใดข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยจึงมีความลำเอียงต่อเพศชายมากกว่าเพศหญิง การที่เกิดเหตุการณ์ดังกล่าวนี้อาจเป็นเพราะว่า

1. โดยปกติทั่วไปแล้ว นักเรียน นิสิต หรือนักศึกษาเพศหญิง เป็นผู้มีความรู้ความสามารถทางภาษาสูงกว่าผู้เรียนเพศชาย ทั้งนี้จากการผลการวิจัยทั้งภายในและต่างประเทศจำนวนมากต่างก็ค้นพบปรากฏการณ์ดังกล่าวแล้ว (เช่น Burstall, 1978: 3-8); ไกรคุง อนุรักษ์กุล, 2525: 54; ประทุมเจริญผล, 2526: 81)

2. ข้อทดสอบจำนวนหนึ่งอาจมีความลำเอียงอันเกิดจากการเลือกเนื้อหาอย่างลำเอียงของผู้สร้างข้อทดสอบอยู่ด้วยการเลือกเนื้อหาที่ไม่เป็นส่วนที่ดีจากหลักสูตรและเนื้อหาที่เลือกมาเอื้อประโยชน์ให้แก่ผู้สอบที่เป็นเพศหญิงมากกว่าเพศชาย รวมทั้งข้อทดสอบอาจมีความลำเอียงที่เกิดจากสร้างข้อทดสอบแต่ละข้อ โดยผู้สร้างอาจใช้คำศัพท์ โครงสร้าง และสำนวนภาษาจำนวนหนึ่งซึ่งเป็นที่รู้จักกันดีเฉพาะผู้สอบเพศหญิง หรือเป็นสิ่งที่ผู้สอบเพศหญิงมีความสนใจมากกว่าผู้สอบเพศชายก็ได้ (Shepard, 1982: 10) ผู้ที่จะตอบข้อสงสัยเหล่านี้ได้จะต้องศึกษาวิเคราะห์ และพิจารณาเนื้อหา ในเชิงตรรก (Logical Analysis) เป็นรายข้ออย่างจริงจัง ประกอบด้วยความรู้ที่ได้จากการวิเคราะห์เชิงสถิติ (Statistic Analysis) ที่ได้จากงานวิจัยเช่นนี้ ทั้งนี้เพราะความลำเอียงของข้อทดสอบเป็นสิ่งที่เกิดขึ้นในข้อทดสอบ

แต่อย่างไรก็ตามเหตุผลทั้ง 2 ประการดังกล่าวแล้วค่อนข้างขัดแย้งกัน และยังไม่อาจเชื่อได้แน่ชัดว่าเหตุผลใดมีน้ำหนักความน่าเชื่อถือได้มากกว่ากัน เนื่องจากแบบทดสอบทั่ว ๆ ไป ที่ใช้ในการเปรียบเทียบผลสัมฤทธิ์ในการเรียนภาษา หรือสมรรถภาพทางภาษา อาจยังไม่ได้คำนึงถึงเรื่องความลำเอียงต่อเพศของข้อทดสอบ เนื่องจากเรื่องความลำเอียงของข้อทดสอบเป็นแนวคิดที่นักทดสอบเพิ่งนำมาปฏิบัติเมื่อไม่นานมานี้ คือเมื่อประมาณ ค.ศ. 1970 นี้เอง (Berk, 1983: 1) ดังนั้นแบบทดสอบที่ใช้ในงานวิจัย เพื่อการเปรียบเทียบดังกล่าวอาจประกอบด้วยข้อทดสอบที่มีความลำเอียงต่อเพศก็ได้ ทำให้ผลการวิจัยยังไม่เชื่อได้มากนัก ดังนั้นข้อค้นพบในอดีตในประเด็นที่ 1 จึงยังเป็นเรื่องที่น่าค้นหาคำตอบต่อไป

2. เหตุใดข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยจึงมีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์อื่นมากกว่าต่อภาคกลาง

ปรากฏการณ์ดังกล่าวนี้อาจเป็นเพราะว่า

1. ความไม่เท่าเทียมกันด้านโอกาสทางการศึกษาของผู้สอบจากภาคกลางและภาคอื่น ๆ ซึ่งโดยมากยังมีสภาพเป็นชนบท ขาดความพร้อมทางอุปกรณ์การเรียน ขาดครู-อาจารย์ผู้สอนที่มีคุณวุฒิทางการศึกษาสูง ขาดโอกาสในการใช้หรือรับรู้สื่อภาษาอังกฤษ รวมทั้งความขาดแคลนอื่น ๆ เมื่อเทียบกับผู้สอบที่อยู่ในภาคกลาง ซึ่งโดยมากอยู่ในสังคมเมือง จึงมีโอกาสดูแลและความพร้อมในด้านต่าง ๆ ดังกล่าวแล้วดีกว่าความแตกต่างดังกล่าวนี้คงเป็นปัจจัยที่สำคัญอย่างหนึ่งที่ทำให้สัมฤทธิ์ผลในการเรียนภาษาอังกฤษของผู้เรียนในต่างจังหวัดต่ำกว่าผู้เรียนในกรุงเทพมหานคร (สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2531) เช่น ในปี พ.ศ.

2516 อ่ำไพ อนันต์ศฤงคาร (2516: 41) ศึกษาพบว่านักเรียนชั้นมัธยมศึกษาปีที่ 3 ในกรุงเทพมหานครมีผลสัมฤทธิ์ในการเรียนภาษาอังกฤษสูงกว่านักเรียนจากจังหวัดสงขลาอย่างมีนัยสำคัญและในปี 2533 เฉลิมชัย มิติภักษ์ (2532: 39-42) ศึกษาพบว่าความสามารถในการฟังภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เรียนอยู่ในโรงเรียนในเขตอำเภอเมืองนครราชสีมา สูงกว่าของนักเรียนที่อยู่นอกเขตอำเภอเมืองอย่างมีนัยสำคัญ เป็นต้น

2. ข้อทดสอบจำนวนหนึ่งอาจมีความลำเอียงอันเกิดจากการเลือกเนื้อหาอย่างลำเอียงของผู้สร้างข้อทดสอบด้วยการเลือกเนื้อหาที่ไม่เป็นสัดส่วนที่ดีจากหลักสูตร และข้อทดสอบอีกจำนวนหนึ่งอาจมีความลำเอียงจากการสร้างข้อทดสอบของกรรมการสร้างข้อทดสอบ โดยผู้สร้างอาจใช้คำศัพท์เนื้อเรื่องสำหรับอ่านโครงสร้าง และสำนวนภาษาจำนวนหนึ่งที่เป็นที่รู้จักกันดีในระหว่างผู้สอบในภาคกลาง แต่ไม่เป็นที่รู้จักกันทั่วไป สำหรับผู้สอบจากภาคอื่น ๆ ความผิดพลาดนี้อาจเกิดจากความไม่ตั้งใจของผู้สร้างข้อทดสอบ แต่สิ่งเหล่านี้ล้วนแต่มีอิทธิพลต่อโครงสร้างความรู้ (schema) ของผู้สอบ และโครงสร้างความรู้ก็มีอิทธิพลต่อความเข้าใจในเรื่องที่อ่านอย่างมีนัยสำคัญด้วย (Mandler, 1979: 263-6; Carrell, 1984: 332-342) ดังนั้นหากผู้สอบอยู่ในสังคมเมืองก็ยังมีโอกาสได้เรียนรู้สิ่งต่าง ๆ รอบตัวมากกว่า ทั้งจากความพร้อมด้านโอกาสทางการศึกษาเชิงรูปนัย (formal education) ที่ดีกว่าแล้ว ก็ยังมีโอกาสทางการศึกษาเชิงอรูปนัย (informal education) เช่นจากสื่อมวลชนต่าง ๆ ดีกว่า

ผู้สอบจากภาคอื่น ๆ ที่มีโอกาสน้อยกว่าด้วย ดังนั้น จากเหตุผลที่สำคัญดังกล่าวแล้ว 2 ข้อข้างต้น อาจเป็นปัจจัยที่สำคัญที่ทำให้แบบทดสอบมีข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคอื่น ๆ มากกว่าจากภาคกลาง

3. เหตุใดจำนวนข้อทดสอบที่ลำเอียงซึ่งได้จากการวิเคราะห์หาความลำเอียงแต่ละวิธีจึงแตกต่างกัน

ความแตกต่างของจำนวนข้อทดสอบที่ลำเอียงดังกล่าวอาจเกิดจากสาเหตุต่อไปนี้คือ

1. แนวคิดเกี่ยวกับความลำเอียงของข้อทดสอบของแต่ละวิธีแตกต่างกัน ทำให้แต่ละวิธีมีนิยามเชิงปฏิบัติการ (operation definition) ของความลำเอียงของข้อทดสอบแตกต่างกัน จึงคำนวณหาความลำเอียงด้วยวิธีการและสูตรที่ต่างกัน จนเป็นเหตุให้จำนวนข้อทดสอบที่ลำเอียงต่างกัน เช่น (Shepard, 1982: 23; Osterlind, 1983: 55):

1. วิธีกำหนดจุดเดลด้า เชื่อว่า

ข้อทดสอบที่ลำเอียงคือข้อทดสอบที่มีค่าเดลด้าของผู้สอบแต่ละกลุ่มเรียงลำดับแตกต่างกัน

2. วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไค-สแควร์ เชื่อว่า

ข้อทดสอบที่ลำเอียงคือข้อทดสอบซึ่งผู้สอนที่ได้คะแนนในช่วงระดับคะแนนเดียวกัน มีสัดส่วนของการตอบข้อทดสอบถูกและ

ผิดแตกต่างกัน

3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ เชื่อว่าข้อทดสอบที่ลำเอียง คือ ข้อทดสอบซึ่งผู้สอบที่มีความสามารถเท่ากันมีโอกาสตอบข้อทดสอบได้ถูกต้องแตกต่างกัน

ดังนั้น จะเห็นได้ว่าวิธีที่ 1 มุ่งพิจารณาที่ระดับความยาก-ง่ายของข้อทดสอบแต่ละข้อ เป็นสำคัญ เนื่องจากคำนวณหาความลำเอียงของข้อทดสอบโดยอาศัยค่าเดลต้า (Δ) ซึ่งเป็นค่าที่แสดงความยาก-ง่ายของข้อทดสอบที่แปลงมาจากค่าอัตราส่วนของการตอบข้อทดสอบแต่ละข้อถูก (p) มากน้อยเพียงใด ซึ่งได้แก่ค่าดัชนีของความยากง่ายของข้อทดสอบโดยตรง

ส่วนวิธีที่ 2 มุ่งพิจารณาที่ระดับความสามารถของกลุ่มผู้สอบและความยากง่ายของข้อทดสอบแต่ละข้อ เป็นสิ่งสำคัญ จึงเป็นวิธีที่ซับซ้อนกว่าวิธีที่ 1 เล็กน้อยในเชิงของแนวคิดและวิธีการคำนวณ แต่ยังเป็นวิธีที่มี จุดอ่อนมากเกี่ยวกับการแบ่งจำนวนกลุ่มของผู้สอบว่าควรจะเป็นกี่กลุ่มสำหรับการวิเคราะห์หาความลำเอียงของข้อทดสอบแต่ละข้อ เนื่องจากยังไม่มีกฎเกณฑ์ที่แน่นอน (Osterlind, 1983: 41-42) และการที่ผู้สอบที่มีคะแนนรวมอยู่ในระดับเดียวกัน อาจมีความสามารถที่แท้จริง แตกต่างกันได้ทั้งนี้ขึ้นอยู่กับช่วงห่างของคะแนน (score interval) และที่มาของคะแนนรวม

สำหรับวิธีที่ 3 นั้นมุ่งพิจารณาความสามารถของผู้สอบแต่ละคน และพารามิเตอร์ของข้อทดสอบแต่ละข้ออย่างละเอียด คือ อำนาจจำแนกของข้อทดสอบ (a) ความยากง่ายของข้อทดสอบ (b) และค่าการเดาถูกของข้อทดสอบ (c) เป็นสำคัญ จึงเป็นวิธีที่

ซับซ้อนมากที่สุดในเรื่องของการคำนวณ และเป็นวิธีที่ไว (sensitive) มากที่สุด ในเรื่องของการค้นหาข้อทดสอบที่ลำเอียง (Osterlind, 1983: 77)

2. เกณฑ์ในการตัดสินความลำเอียงของข้อทดสอบแต่ละวิธีไม่แน่นอน เนื่องจากยังขาดเกณฑ์กลางที่เป็นที่ยอมรับกันทั่วไป ดังนั้นงานวิจัยแต่ละเรื่อง จึงอาจใช้เกณฑ์ที่แตกต่างกัน เช่น เกณฑ์ในการตัดสินระยะห่างตั้งฉากของค่าอันดับคู่เดลต้าไปยังแกนสำคัญ (d) นักวิจัยอาจใช้ $d = 2S_d, 3S_d, 0.75 z\text{-score unit}$ หรือ $1.5 z\text{-score unit}$ ก็ได้ (Angoff, 1982: 107; Osterlind, 1983: 35) ทั้งนี้แล้วแต่วัตถุประสงค์ของการวิจัยแต่ละเรื่อง หรือในกรณีของการวิเคราะห์วิธีที่ 2 (Chi-square Method) ผู้วิจัยอาจกำหนดระดับความมีนัยสำคัญของตนเองได้ว่าควรจะเป็น $\alpha = 0.05$ หรือ 0.01 รวมทั้งการกำหนดจำนวนกลุ่มของผู้สอบซึ่งได้กล่าวมาแล้วว่ายังไม่มีความแน่นอน จำนวนกลุ่มของผู้สอบที่แบ่งตามระดับความสามารถโดยใช้คะแนนรวมนี้ มีผลโดยตรงต่อเกณฑ์ในการตัดสินความลำเอียงของข้อทดสอบด้วย เนื่องจากสูตรทั่วไปในการคำนวณหาค่าระดับความมีนัยสำคัญ เมื่อคำนึงถึงค่าอัตราความคลาดเคลื่อน (error rate) แล้ว คือ $\alpha = \alpha_j$ และ j คือจำนวนกลุ่ม ดังนั้น ถ้าหากว่าจำนวนกลุ่มเปลี่ยนแปลงไป เกณฑ์ในการตัดสินความลำเอียงของข้อทดสอบก็เปลี่ยนแปลงไปด้วย ส่วนวิธีวิเคราะห์ที่ 3 คือ Three-parameter Logistic Method นั้น แม้ว่ามีนักวิจัยบางท่านเสนอว่าค่า $0 < 0.40 =$ ลำเอียงน้อย 0 ระหว่าง $0.40-0.70 =$ ลำเอียงปานกลาง และ $0 > 0.70 =$ ลำเอียงมาก (Rudner, 1977 อ้างจากสุรศักดิ์อมรรตน์ศักดิ์, 2531: 44) แต่ไม่ได้ระบุไว้แน่ชัดว่าคำว่า 'ลำเอียงน้อย' นั้นมี 'นัยสำคัญ' หรือไม่ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงไม่ได้ค้น

หาข้อทดสอบที่มีความลำเอียงน้อย

ดังนั้นจะเห็นได้ว่าเกณฑ์ในการตัดสินความลำเอียงของข้อทดสอบของการวิเคราะห์แต่ละวิธียังไม่เกณฑ์กลางที่ยอมรับกันทั่วไป หากเปลี่ยนแปลงเกณฑ์ของวิธีใดวิธีหนึ่งก็ย่อมทำให้จำนวนข้อทดสอบที่ลำเอียงเปลี่ยนแปลงตามไปด้วยด้วยเหตุผลที่สำคัญดังกล่าวแล้ว

2 ประการข้างต้น น่าจะเป็นสาเหตุที่ทำให้ข้อทดสอบที่ลำเอียงแต่ละวิธีแตกต่างกัน และเมื่อพิจารณาถึงความซับซ้อนของแนวคิดเกี่ยวกับความลำเอียงของข้อทดสอบและวิธีการคำนวณแล้ว อาจกล่าวได้ว่าวิธีที่ 1 เป็นวิธีที่อนุรักษ์นิยม (conservative) มากที่สุด รองลงมาได้แก่ วิธีที่ 2 และวิธีที่ 3 เป็นวิธีที่ไว (sensitive) มากที่สุด เมื่อคำนึงถึงการค้นหาข้อทดสอบที่ลำเอียงในระดับน้อย หรือต่ำด้วย ดังนั้นโดยสภาพทั่วไปแล้วจำนวนข้อทดสอบที่ลำเอียงที่พบจากวิธีที่ 3 จึงมีจำนวนมากที่สุด (แม้ไม่ได้รวมข้อที่ลำเอียงน้อยด้วย) รองลงมาได้แก่ วิธีที่ 2 และวิธีที่ 1 พบข้อทดสอบที่ลำเอียงจำนวนน้อยที่สุด ผลการวิจัยครั้งนี้จึงสอดคล้องกับสภาพทั่วไปดังกล่าวแล้ว และสอดคล้องกับผลการวิจัยของทัศนีย์ พิรมนตรี (2530, 73) ที่พบว่าข้อทดสอบที่ลำเอียงต่อภาคภูมิศาสตร์ของวิชาคณิตศาสตร์ที่วิเคราะห์โดยวิธีที่ 3 มีมากกว่าวิธีที่ 1 (ส่วนวิธี (Chi-square Method) นั้น ทัศนีย์ พิรมนตรี ใช้วิธีที่ต่างจากวิธีที่ผู้วิจัยใช้ในการวิจัยครั้งนี้) และสอดคล้องกับผลการวิจัยของสุรศักดิ์ อมรรัตน์-ศักดิ์ (2531: 170-182) ในการเปรียบเทียบผลของการวิเคราะห์หาความลำเอียงของ 4 วิธีจากการวิเคราะห์แบบทดสอบคัดเลือก 4 ฉบับ

4. เหตุใดจำนวนข้อทดสอบที่ลำเอียงจากการวิเคราะห์แต่ละวิธีจึงมีความสัมพันธ์กัน อย่างไม่มีนัยสำคัญ

ปรากฏการณ์ดังกล่าวนี้อาจเป็นเพราะว่า

1. การวิเคราะห์หาความลำเอียงแต่ละวิธีมุ่งวิเคราะห์ลักษณะของข้อทดสอบข้อเดียวกันคนละด้าน เนื่องจากมีแนวคิดเกี่ยวกับความลำเอียงแตกต่างกัน และวิธีการทดสอบแตกต่างกัน ตามรายละเอียดที่ได้กล่าวมาแล้วข้างต้นในประเด็นที่ 3
2. วิธีการวิเคราะห์หาความลำเอียงทั้ง 3 วิธียังมีข้อบกพร่องบางอย่างเช่น
 - ก. ข้อทดสอบที่ลำเอียงที่พบจากวิธีที่ 1 มักเป็นข้อที่มีอำนาจจำแนกสูงหรือต่ำมาก นอกจากเป็นข้อทดสอบที่มีค่าความยากง่าย (Δ) ของกลุ่มแตกต่างกันมาก ดังนั้น หากข้อทดสอบมีค่าอำนาจของกลุ่มแตกต่างกันมาก ค่าความลำเอียงของข้อทดสอบอาจเป็นค่าที่ผิดพลาด (Angoff, 1982: 104)
 - ข. ข้อทดสอบที่ลำเอียงที่พบจากวิธีที่ 2 ได้รับอิทธิพลมาจากลักษณะของการกระจายของคะแนนสอบ (score distribution) ของกลุ่มที่นำมาเปรียบเทียบด้วย ดังนั้นข้อทดสอบจะถือว่ามีลักษณะลำเอียงหรือไม่ ยิงขึ้นอยู่กับลักษณะการกระจายของคะแนนสอบด้วย ซึ่งเป็นปัจจัยที่อยู่นอกเหนือค่านิยามของความลำเอียงของวิธีนี้ (Ironson, 1982: 136)
 - ค. มีโอกาสที่เป็นไปได้ที่ข้อทดสอบมีความลำเอียงแต่วิธีการวิเคราะห์วิธีที่ 3 ไม่อาจค้นหาพบได้ เนื่องจากข้อทดสอบที่ค่าพารามิเตอร์ต่างกันอาจมีโค้งลักษณะของข้อทดสอบ (ICC) เหมือนกันก็ได้ (Ironson, 1982: 149)

ดังนั้น จากข้อบกพร่องทั้ง 3 ข้างต้นนี้ จึงทำให้กล่าวได้ว่าจำนวนข้อทดสอบที่ลำเอียงนั้นอาจมีบางข้อที่ไม่ใช่ข้อที่ลำเอียง และอาจมีบางข้อที่วิธีวิเคราะห์ทั้ง 3 ไม่อาจค้นพบได้ และจากการที่จำนวนข้อทดสอบที่ลำเอียงที่เพิ่มขึ้นหรือลดลงอย่างไม่มีระบบ รวมทั้งการค้นหาค่าความลำเอียงของแต่ละวิธี มุ่งวิเคราะห์ลักษณะของข้อทดสอบคนละด้าน จึงอาจเป็นเหตุให้จำนวนข้อทดสอบที่ลำเอียงของแต่ละวิธีมีความสัมพันธ์กันอย่างไม่มีความสำคัญ ผลการวิจัยครั้งนี้ขัดแย้งกับผลการวิจัยของ Rudner และคณะ (Rudner and Others, 1980:8) ซึ่งศึกษาเปรียบเทียบผลของการวิเคราะห์หาข้อทดสอบที่ลำเอียง 7 วิธีโดยอาศัยข้อมูลที่สร้างขึ้นจากวิธี Monte Carlo และพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของจำนวนข้อทดสอบที่พบจาก 3 วิธีที่ผู้วิจัยศึกษาครั้งนี้มีค่าในระดับปานกลางคือ r_{xy} ระหว่าง 0.60-0.66 ซึ่ง Rudner และคณะกล่าวว่าค่าสูงกว่าผลการวิจัยที่พบโดย Ironson (1979) และ Rudner และ Coney (1978) ซึ่งศึกษาจากข้อมูลจริงที่มีขนาดน้อยกว่า (อ้างจาก Rudner และคณะ, 1980: 9) แต่ข้อค้นพบที่ขัดแย้งกันนี้อาจเป็นไปได้ เพราะการศึกษาครั้งนี้ผู้วิจัยอาศัยข้อมูลเชิงประจักษ์ (empirical data) จึงมีจำนวนการกระทำซ้ำ ในการวิเคราะห์น้อยครั้งกว่าการวิจัยที่ Rudner และคณะจัดกระทำขึ้นกับข้อมูลจำลอง ซึ่งกระทำซ้ำได้จำนวนมากมายและภายใต้หลายเงื่อนไข

ข้อเสนอแนะ

- ก. สำหรับคณะกรรมการสร้างแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยและผู้ที่เกี่ยวข้อง

ควรศึกษาและวิเคราะห์ข้อทดสอบที่ผลการวิจัยครั้งนี้พบว่ามีความลำเอียงในลักษณะต่าง ๆ ในเชิงของการวิเคราะห์เนื้อหาและตรรก เพื่อค้นหาว่าเหตุใดข้อทดสอบเหล่านั้นจึงมีความลำเอียงต่อเพศหรือต่อภาคภูมิศาสตร์ของผู้สอบ ทั้งนี้เพื่อให้ได้แนวทางในการเลือกเนื้อหาและการสร้างข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยต่อไป เพื่อให้ข้อทดสอบมีความเป็นธรรมแก่ผู้สอบมากที่สุด เพราะ 'การตัดสินใจให้ผู้สมควรสอบได้ตกและตัดสินใจให้ผู้สมควรสอบตกได้' เป็นความผิดพลาดที่ร้ายแรงที่คณะกรรมการสร้างข้อทดสอบควรตระหนักไว้เสมอ และควรพยายามทุกวิถีทางให้เกิดเหตุการณ์ดังกล่าวให้น้อยที่สุด เพราะการตัดสินใจที่ผิดพลาดดังกล่าวมีผลกระทบต่ออนาคตของบุคคลทั้งผู้ที่สอบได้และสอบตกและคุณภาพของประชากรของประเทศเป็นอย่างมาก

ข. สำหรับนักวิจัย

ควรศึกษาและวิเคราะห์ข้อทดสอบที่ผลการวิจัยครั้งนี้พบว่ามีความลำเอียง ในเชิงของการวิเคราะห์เนื้อหาและตรรกว่าเป็นเพราะเหตุใดสอดคล้องกับความเป็นจริงมากน้อยเพียงใด และวิธีวิเคราะห์ใดที่ให้ผลสอดคล้องกับความเป็นจริงมากที่สุด

ค. สำหรับผู้สร้างและพัฒนาแบบทดสอบทั่วไป

1. เนื่องจากมีผลการวิจัยจำนวนมากค้นพบและยืนยันว่า การวิเคราะห์หาความลำเอียงที่ได้ผลดีที่สุด คือ Three-parameter Logistic Method [เพราะละเอียดมากที่สุดให้รายละเอียดได้มากที่สุดและสอดคล้องกับแนวคิดทางการทดสอบแนวใหม่มากที่สุด] รองลงมาได้แก่ Chi-square Method ที่แบ่งผู้สอบออกเป็นกลุ่มย่อย 5 กลุ่ม (Subkoviak and Others, 1984: 56; Rudner and Others,

1980: 9; Ironson. 1979 อ้างจาก Burrill, 1982: 169) ดังนั้น หากเป็นไปได้ควรวิเคราะห์หาความลำเอียงของข้อทดสอบ โดย Three-parameter Logistic Method แต่วิธีดังกล่าวนี้เสียค่าใช้จ่ายและเวลาในการคำนวณมากที่สุดรวมทั้งต้องอาศัยผู้สอบมากกว่า 1,000 คน และโปรแกรมคอมพิวเตอร์ด้วย

ดังนั้น หากไม่สามารถกระทำได้ในทางปฏิบัติก็ควรใช้ Chi-square Method แทน หรือหากยังไม่สามารถกระทำได้ ก็ควรใช้ Delta

Plot Method ซึ่งเป็นวิธีที่ง่ายที่สุด และสะดวกที่สุดในทางปฏิบัติ

2. ข้อทดสอบที่ลำเอียงไม่จำเป็นที่จะต้องเป็นข้อทดสอบที่ 'ไม่ดี' เสมอไปเพราะในบางสถานการณ์ ข้อทดสอบที่ลำเอียงอาจให้ข้อสารสนเทศที่ดีที่สุด หรือเป็นตัวทำนายที่ดีที่สุด เช่น ในการเลือกวิชาเรียน (Rudner and Others, 1984 b: 227) ดังนั้นในการพัฒนาข้อทดสอบควรคำนึงถึงเรื่องวัตถุประสงค์ของแบบทดสอบเป็นสำคัญด้วยเสมอ

เอกสารอ้างอิง

- ไกรคุง อ้นคชกุล. 'การศึกษาผลสัมฤทธิ์ในการเรียนโครงสร้างไวยากรณ์อังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในกรุงเทพมหานคร.' *วิทยานิพนธ์การศึกษามหาบัณฑิต*. กทม.: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2525.
- คณะกรรมการการศึกษาแห่งชาติ. *การวิจัยและประเมินคุณภาพโรงเรียนประถมศึกษา*. กทม. : สำนักงานคณะกรรมการการศึกษาแห่งชาติ สำนักนายกรัฐมนตรี , 2531.
- เฉลิมชัย มิตินักษ์. 'การศึกษาความสามารถในการฟังภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนมัธยมศึกษาที่ตั้งอยู่ในเขตอำเภอเมืองและโรงเรียนที่ตั้งอยู่นอกเขตอำเภอเมือง.' *วิทยานิพนธ์การศึกษามหาบัณฑิต*. กทม. : บัณฑิตวิทยาลัย มหาวิทยาลัย เกษตรศาสตร์, 2522.
- ประทุม เจริญผล. 'องค์ประกอบบางประการที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษของนักศึกษาผู้ใหญ่สายสามัญ ระดับที่ 5 ในเขตการศึกษา 5.' *วิทยานิพนธ์การศึกษามหาบัณฑิต*. นครปฐม: บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, 2526.
- ฝ่ายสารสนเทศ 4. 'สถิติการศึกษา.' *วารสารการศึกษาแห่งชาติ*. 2 (ธันวาคม, 2529 - มกราคม 2530): 45-50.
- ทัศนีย์ พิรมนตรี. 'การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526.' *วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต*. กทม.: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2530.

- สุพัฒน์ สุขมลสันต์. 'ความสอดคล้องของข้อทดสอบคัดเลือกเข้ามหาวิทยาลัยกับเกณฑ์ในทฤษฎีการวัดผล.' ใน *ความสอดคล้องของข้อทดสอบเข้ามหาวิทยาลัยกับหลักสูตรภาษาอังกฤษ และแนวโน้มในการเรียนการสอนภาษาอังกฤษ เพื่อการสื่อสารและการสร้างคลังข้อสอบ*. กทม.: Thai-TESOL's Occasional Papers No. 1984/1, 1984.
- สุพัฒน์ สุขมลสันต์. 'ความสอดคล้องของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี 2528 กับเกณฑ์ในทฤษฎีการวัดผล.' บทความเสนอในการประชุมใหญ่ของสมาคม Thai/TESOL เรื่อง *การวิเคราะห์ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย*, 2528.
- สุรศักดิ์ อมรรัตนศักดิ์. 'การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี.' *วิทยานิพนธ์ครุศาสตร์ดุสิตบัณฑิต*. กทม.: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2531.
- อำไพ อนันต์ศฤงคาร. 'สัมฤทธิ์ผลทางการเรียนวิชาภาษาอังกฤษชั้นมัธยมศึกษาปีที่ 3 โรงเรียนมหาวิทยาลัยราชภัฏ.' *วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต*. กทม.: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2516.
- Angoff, W.H. 'Use of Difficulty and Discrimination Indices for Detecting Item Bias.' In *Handbook of Methods for Detecting Test Bias*. R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Berk, R.A. *Handbook of Methods for Detecting Test Bias*. Maryland: The Johns Hopkins University Press, 1982.
- Burstall, C. 'Factors Affecting Foreign Language Learning : A Consideration of Some Recent Research Findings.' In *Language Teaching and Linguistics*. V. Kinsella (ed.) London: Oxford University Press, 1978.
- Burrill, L.E. 'Comparative Studies of Item Bias Methods.' In *Handbook of Methods for Detecting Test Bias*. R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Bradford, A. 'A Farang's Problems with the University Entrance Examinations in English.' ใน *การวิเคราะห์ข้อทดสอบเข้ามหาวิทยาลัยวิชาภาษาอังกฤษ*. Priya Unaratana (ed.) กทม. สมาคม Thai/TESOL Occasional Papers No. 1991/1, 1990.
- Carrell, P.L. 'Schema Theory and ESL Reading: Classroom Implications and Applications.' *The Modern Language Journal*. 68 (1984): 332-342.
- Ironson, G.H. 'Use of Chi-square and Latent Trait Approaches for Detecting Items Bias.' In *Handbook of Methods for Detecting Test Bias*, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Ironson, G.H. and Subkoviak, M.J. 'A Comparison of Several Methods of Assessing Item Bias.' *Journal of Educational Measurement* 16 (1979): 209-225.

- Mandler, J.M. 'Categorial and Schematic Organization in Memory.' In *Memory Organization and Structure*. C.R. Puff (ed.). New York: Academic Press, 1979.
- Osterlind, S.J. *Test Item Bias*. California: SAGE Publications. Inc, 1983.
- Rudner, L.M. and Others. 'A Monte Carlo comparision of Seven Biased Item Detection Techniques.' *Journal of Educational Measurement*. 17 (Spring 1980): 1-10.
- Shepard. L.A. 'Definitions of Bias.' In *Handbook of Methods for Detecting Test Bias*. R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Subkoviak. M.J. and Others. 'Empirical Comparison of Selected Item Bias Detection Procedures with Bias Manipulation.' *Journal of Educational Measurement*. 21 (Spring 1984): 49-58.