

## บทที่ 4

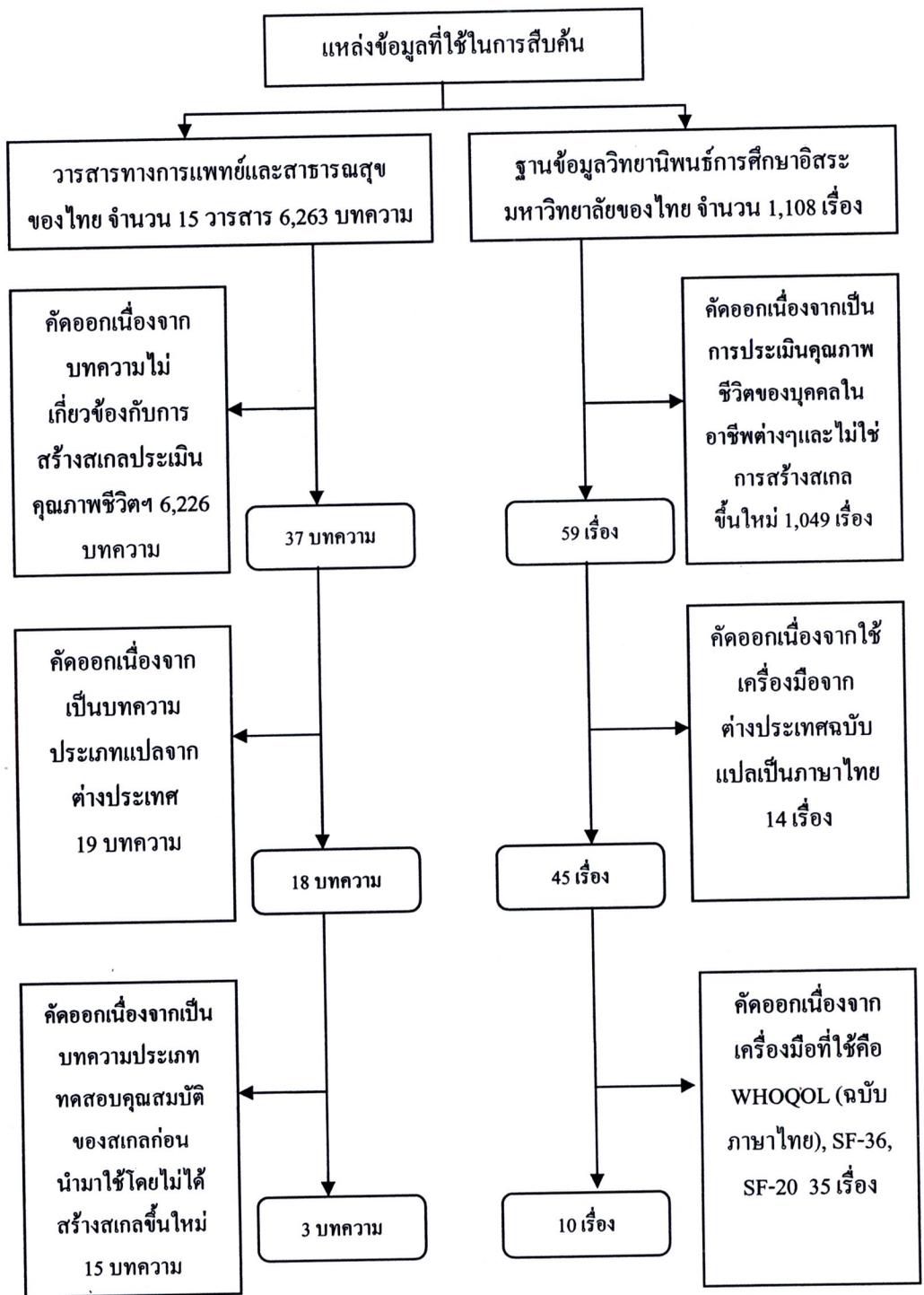
### ผลการวิจัยและการอภิปรายผล

จากการเก็บรวบรวมข้อมูลบทความวิจัยทุกบทความที่พิมพ์เผยแพร่ ในวารสารทางการแพทย์และสาธารณสุขของไทย 15 วารสาร จำนวนทั้งหมด 6,263 บทความ และวิทยานิพนธ์ระดับบัณฑิตศึกษา ทางการแพทย์และสาธารณสุขของมหาวิทยาลัยในประเทศไทย จำนวน 1,108 เรื่อง ที่พิมพ์เผยแพร่ในช่วงเวลาระหว่างเดือนมกราคม พ.ศ.2543 ถึง เดือนธันวาคม พ.ศ.2552

ผลการสืบค้นพบว่า บทความวิจัยที่เผยแพร่ในวารสารการแพทย์และสาธารณสุขของไทย จำนวน 3 บทความ และข้อมูลวิทยานิพนธ์ระดับบัณฑิตศึกษา ทางการแพทย์และสาธารณสุขของมหาวิทยาลัยในประเทศไทย จำนวน 10 เรื่อง ที่เป็นไปตามเกณฑ์การคัดเลือกที่กำหนดไว้ ดังแสดงในภาพที่ 3

รายละเอียดระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้ของงานที่ทบทวนทั้ง 13 งาน ในประเด็นต่าง ๆ ที่พิจารณาตามมาตรฐานของการสร้างสเกล 11 ประเด็น ปรากฏดังตาราง 3 ภาคผนวก ก

จากนั้นผู้วิจัยพิจารณาการรายงานระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างสเกลวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ตามประเด็นที่กำหนดไว้ แล้วนำเสนอเป็นผลการวิจัยและอภิปรายผล ตามลำดับดังนี้



ภาพที่ 3 ผลการสืบค้นบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษา จากแหล่งต่าง ๆ



## 1. ผลการวิจัย

### 1.1 ลักษณะทั่วไปของบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษา

บทความวิจัยและวิทยานิพนธ์ ด้านการแพทย์และสาธารณสุขที่นำมาศึกษา มีทั้งหมด 13 เรื่อง พบว่า 9/13 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อสร้างสเกลวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ และ 4/13 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อประเมินคุณภาพชีวิตแต่ผู้วิจัยสร้างสเกลวัดคุณภาพชีวิตขึ้นใหม่ ซึ่งพบว่า 9/13 เรื่อง สร้างสเกลเพื่อวัดตัววัดสุขภาพเกี่ยวกับโรคเรื้อรัง เช่น คุณภาพชีวิตผู้มีปัญหาสายตา คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด ที่เหลือเป็นตัววัดสุขภาพในประชากรเฉพาะกลุ่ม ได้แก่ คุณภาพชีวิตสตรีวัยหมดประจำเดือน ผู้เลิกบุหรี่ และผู้ป่วยเด็ก รายละเอียดอื่นๆ ของงานวิจัยที่ศึกษาปรากฏดังตารางที่ 2

ตารางที่ 2 ลักษณะทั่วไปของบทความวิจัยและวิทยานิพนธ์ที่ศึกษา

ลักษณะทั่วไป	บทความ	วพ. บัณฑิตศึกษา	รวม
	(n=3) จำนวน	(n=10) จำนวน	(n=13) จำนวน
<b>วัตถุประสงค์การศึกษา</b>			
- มีวัตถุประสงค์หลักเพื่อสร้างสเกลเพื่อวัดคุณภาพชีวิต	3	6	9
- มีวัตถุประสงค์หลักเพื่อประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ แต่สร้างสเกลขึ้นใหม่	-	4	4
<b>ตัววัดสุขภาพและประชากรเป้าหมาย</b>			
- คุณภาพชีวิตผู้มีปัญหาด้านสายตา	1	1	2
- คุณภาพชีวิตผู้มีปัญหาทางจิต	1	-	1
- คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ	-	1	1
- คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด	-	1	1
- คุณภาพชีวิตผู้ป่วยหลังผ่าตัดเปิดกะโหลก	-	1	1
- คุณภาพชีวิตผู้ป่วยมะเร็งระบบสืบพันธุ์	-	1	1
- คุณภาพชีวิตผู้ป่วยเบาหวาน	-	1	1
- คุณภาพชีวิตผู้ป่วยไตวายเรื้อรัง	-	1	1
- คุณภาพชีวิตหญิงวัยหมดประจำเดือนที่มีภาวะกระดูกพรุน	-	1	1
- คุณภาพชีวิต ในคนไทยที่เลิกบุหรี่	-	1	1
- คุณภาพชีวิต ในผู้ป่วยเด็ก	-	1	1
- คุณภาพชีวิต ในผู้ให้บริการทางเพศ	1	-	1

## 1.2 การรายงานระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างสเกลตามประเด็นที่กำหนด

ฯ

งานวิจัยที่ทบทวนมีทั้งหมด 13 เรื่อง เป็นบทความวิจัย 3 บทความ วิทยานิพนธ์ 10 เรื่อง เมื่อพิจารณาระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้โดยอิงตามมาตรฐานการสร้างสเกล พบว่า ขั้นตอนหรือประเด็นที่มีการรายงานมากที่สุดคือพบในทุกงานวิจัย (13/13 เรื่อง) คือ วิธีการนิยามตัวแปรที่จะวัด การรวบรวมคำถาม การกำหนดประเภทและรูปแบบสเกล และการประเมินความเที่ยงของสเกลโดยใช้สัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ประเด็นหรือขั้นตอนอื่น ๆ ที่มีการรายงาน เมื่อเรียงลำดับจากพบมากไปน้อย เป็นดังนี้ การทดสอบสเกลเบื้องต้น พร้อมให้รายละเอียดของวิธีการทดสอบและขนาดตัวอย่าง พบ 12/13 เรื่อง การตรวจสอบความตรงตามเนื้อหา พบ 11/13 เรื่อง โดย 7/11 เรื่องประเมินโดยผู้เชี่ยวชาญ 3/11 เรื่องเสนอค่าดัชนีความตรงเชิงเนื้อหา (content validity index; CVI) การวิเคราะห์รายการคำถาม พบ 10/13 เรื่อง โดย 8/9 เรื่องใช้วิธีการวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis) การนำไปใช้ พบ 6/13 เรื่อง โดยรายงานการกำหนดค่าคะแนนและการแปลผล การตรวจสอบความตรงตามโครงสร้าง พบ 5/13 เรื่อง โดยสถิติที่ใช้ เช่น สหสัมพันธ์อย่างง่าย Mann-Whitney U test, Independent two sample t-test และการวิเคราะห์ปัจจัยเชิงยืนยัน (confirmatory factor analysis) การตรวจสอบความตรงตามเกณฑ์สัมพัทธ์ด้วยสหสัมพันธ์อย่างง่าย พบ 2/13 เรื่อง และปรากฏว่าในทุกงานไม่พบการรายงานการประเมินความไว (sensitivity to change) รายละเอียดดังนี้

1.2.1 การกำหนดแนวคิด คำนิยามหรือโครงสร้างของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง รายงานว่าทำขั้นตอนนี้ โดยมี 1 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต มี 9 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา มี 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 1 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต

เมื่อจำแนกตามบทความวิจัย ที่ทบทวนทั้งหมด 3 เรื่อง มี 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา และมี 1 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย

เมื่อจำแนกตามวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 10 เรื่อง มี 1 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต มี 7 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา ที่เหลือ 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิตและการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย วิธีการใดวิธีการหนึ่งร่วมกันอย่างน้อย 2 วิธี

1.2.2 การรวบรวมรายการคำถาม เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง ทุกงานรายงานว่าทำขั้นตอนนี้ โดยมี 2 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และ/หรือสัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 3 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย มี 8 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา

เมื่อจำแนกตามบทความวิจัย ที่ทบทวนทั้งหมด 3 เรื่อง มี 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา ที่เหลือ 1 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย

เมื่อจำแนกตามวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 10 เรื่อง มี 2 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และ/หรือสัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 6 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา

1.2.3 การกำหนดประเภทและรูปแบบของสเกล เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า ทั้งหมดใช้สเกลของลิเคิร์ต (Likert scale) โดย 11/13 เรื่อง กำหนดตัวเลือกของคำตอบเป็น 5 ระดับ ที่เหลืออีก 2 เรื่องกำหนดตัวเลือกของคำตอบเป็น 3 ระดับ

เมื่อจำแนกตามบทความวิจัย พบว่าที่ทบทวนทั้งหมด 3 เรื่อง พบว่า ทั้งหมดใช้สเกลของลิเคิร์ต (Likert scale) โดย 3/3 เรื่อง กำหนดตัวเลือกของคำตอบเป็น 5 ระดับ เช่นเดียวกับวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 10 เรื่อง พบว่าทั้งหมดใช้สเกลของลิเคิร์ต (Likert scale) โดย 8/10 เรื่อง กำหนดตัวเลือกของคำตอบเป็น 5 ระดับ

1.2.4 การประเมินความตรงตามเนื้อหา จากการศึกษาครั้งนี้พบว่า มี 2/13 เรื่อง ไม่รายงานการประเมินความตรงด้านเนื้อหา ที่เหลือ 11/13 รายงานว่าประเมินความตรงด้านเนื้อหา แต่

รายละเอียดของวิธีการที่ใช้แตกต่างกัน โดยพบดังนี้ 3/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยวิธีการนำไปทดสอบในกลุ่มเป้าหมาย ทำได้โดยนำเครื่องมือที่สร้างขึ้น ไปสัมภาษณ์กับกลุ่มเป้าหมายโดยตรง พิจารณาความเข้าใจในการตอบข้อคำถาม หากกลุ่มเป้าหมายไม่เข้าใจข้อความใดจะตัดข้อคำถามนั้นออกไป และการเลือกคำตอบจากตัวเลือกในแต่ละข้อคำถามถ้ามีการเลือกคำตอบในข้อใดข้อหนึ่งมากเกินไปหรือละ 90 จะตัดข้อคำถามข้อนั้นออกไป 8/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยใช้วิธีขอความเห็นจากผู้เชี่ยวชาญ กรณีเน้นวิธีการเชิงคุณภาพคือให้ผู้เชี่ยวชาญให้ความเห็นในส่วนต่าง ๆ ของเครื่องมือ แล้วผู้วิจัยนำความเห็นของผู้เชี่ยวชาญมาปรับแก้เครื่องมือ สำหรับจำนวนผู้เชี่ยวชาญที่ใช้ในการตรวจสอบความตรงด้านเนื้อหา พบว่ามีจำนวนระหว่าง 3-9 คน โดยส่วนใหญ่ใช้ผู้เชี่ยวชาญ 5 คน

เมื่อจำแนกตามบทความวิจัย พบว่ามีการรายงานความตรงตามเนื้อหา จำนวน 1/3 เรื่อง โดยใช้วิธีการตรวจสอบเชิงคุณภาพโดยผู้เชี่ยวชาญ สำหรับวิทยานิพนธ์ระดับบัณฑิตศึกษา พบว่ามีรายงานการประเมินความตรงตามเนื้อหา จำนวน 10 เรื่อง ในจำนวนนี้ใช้วิธีการตรวจสอบเชิงคุณภาพโดยผู้เชี่ยวชาญ จำนวน 7 เรื่อง ใช้การตรวจสอบเชิงปริมาณโดยหาค่าดัชนีความตรงเชิงเนื้อหา (CVI) และตรวจสอบเชิงปริมาณโดยนำไปทดสอบในกลุ่มเป้าหมาย 3 เรื่องเท่านั้น

1.2.5 การทดสอบสเกลเบื้องต้น จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบ 12/13 เรื่อง รายงานการทดสอบสเกลเบื้องต้น โดยนำไปทดสอบกับกลุ่มคนที่มีลักษณะเหมือนประชากรเป้าหมาย จาก 12 เรื่องของงานที่ทบทวนในการวิจัยครั้งนี้ ที่รายงานการทดสอบสเกลเบื้องต้น มี 5/12 เรื่องใช้การสัมภาษณ์จากตัวอย่างที่มีลักษณะเหมือนประชากรเป้าหมาย ขนาดตัวอย่างที่ใช้มีจำนวน 30, 47, 52 และ 60 คน ตามลำดับ และ 3/12 เรื่อง ใช้การตอบด้วยตนเอง ขนาดตัวอย่างที่ใช้มีจำนวน 6, 20, 30 คน ตามลำดับ ส่วนงานที่เหลือไม่ระบุวิธีการ

เมื่อจำแนกตามบทความวิจัย ทบทวนทั้งหมด 3 เรื่อง พบ 2 เรื่อง รายงานการทดสอบสเกลเบื้องต้น โดยนำไปทดสอบกับกลุ่มคนที่มีลักษณะเหมือนประชากรเป้าหมาย เช่นเดียวกันกับวิทยานิพนธ์ที่ทบทวนทั้งหมด 10 เรื่อง ซึ่งทั้งหมดรายงานการทดสอบสเกลเบื้องต้น โดยนำไปทดสอบกับกลุ่มคนที่มีลักษณะเหมือนประชากรเป้าหมายโดยวิธีการสัมภาษณ์และตอบด้วยตนเอง

1.2.6 การวิเคราะห์รายการคำถาม ในการศึกษาครั้งนี้ เมื่อพิจารณาบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า 4/13 เรื่อง ไม่รายงาน ที่เหลือ 9/13 เรื่องรายงาน ใน 9 เรื่องที่รายงาน มีรายละเอียดดังนี้ 1 เรื่อง วิเคราะห์รายการคำถามทุกวิธี ได้แก่ 1) การพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม (distribution of response) 2) วิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม (inter-item correlation) 3) วิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อ

ค่าตามกับคะแนนรวม (item-total correlation) 4) วิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถาม (inter-item consistency) และ 5) การวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis) เพื่อหาโครงสร้างของข้อมูล

พบว่า มี 2 เรื่อง ใช้วิธีวิเคราะห์รายการคำถามทุกวิธีดังเสนอข้างต้น ยกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม โดยสถิติที่ประเมินความสัมพันธ์คือสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และสถิติที่ใช้ประเมินความสอดคล้องภายในคือสัมประสิทธิ์แอลฟาของครอนบาค 1 เรื่อง ใช้วิธีวิเคราะห์รายการคำถามทุกวิธีดังเสนอข้างต้น ยกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม 1 เรื่อง ใช้วิธีการพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม และการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล 1 เรื่อง ใช้วิธีการวิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถามและการวิเคราะห์ปัจจัยเชิงสำรวจ 2 เรื่อง ใช้การวิเคราะห์ปัจจัยเชิงโครงสร้างเพื่อหาโครงสร้างของข้อมูล 1 เรื่อง ไม่รายงานวิธีการที่ใช้

จะเห็นว่ากรณีการวิเคราะห์ปัจจัยเชิงสำรวจพบ 8/9 เรื่องที่ใช้วิธีการนี้ มีรายละเอียดของการประเมินดังนี้ ทุกงานใช้การสกัดปัจจัยด้วยวิธีการวิเคราะห์องค์ประกอบ (principle component analysis) มีการคัดเลือกจำนวนปัจจัยที่เหมาะสมโดยพิจารณาจากค่า eigenvalue ที่มากกว่า 1, ในจำนวนนี้เลือกวิธีการหมุนแกนแบบมุมแหลม (oblique) จำนวน 2 เรื่อง ด้วยวิธี promax ที่เหลือจำนวน 6 เรื่อง ใช้วิธีการหมุนแกนแบบมุมฉาก (orthogonal) ด้วยวิธี varimax 5 เรื่อง และวิธี quatimax 1 เรื่อง, สำหรับเกณฑ์ในการคัดเลือกข้อคำถามในแต่ละปัจจัยจะมีเกณฑ์พิจารณาจากค่า factor loading ที่มีค่ามากกว่า 0.4 ข้อคำถามนั้นจะคงอยู่ในปัจจัยนั้น จากงานที่ทบทวนพบ 7 งานที่ใช้เกณฑ์นี้ และที่เหลือ 1 งานใช้เกณฑ์จากค่า factor loading ที่มีค่า 0.7 ขึ้นไป

นอกจากนี้พบว่า 5/9 เรื่อง ใช้การวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล ร่วมกับการวิเคราะห์ความสอดคล้องภายใน

เมื่อจำแนกตามบทความวิจัย ที่ทบทวนทั้งหมด 3 เรื่อง พบว่าทั้ง 3 เรื่อง รายงานการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล และอีก 1 เรื่อง รายงานวิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถามด้วยวิธีการสัมประสิทธิ์แอลฟา

สำหรับวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 9 เรื่อง พบว่า 4/9 เรื่อง ไม่รายงาน ที่เหลือ 6/9 เรื่อง รายงาน รายละเอียดดังนี้ 1 เรื่อง รายงานครบทุกวิธีดังกล่าวข้างต้น และมี 2 เรื่อง รายงานครบทุกวิธี ยกเว้น วิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม อีก 1 เรื่อง รายงานครบทุกวิธี ยกเว้น วิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการ

คำถาม 1 เรื่อง เป็นพิจารณาการกระจายคำตอบของแต่ละข้อคำถามและการวิเคราะห์ปัจจัยเชิงสำรวจ สดท้ายอีก 1 เรื่องพิจารณาการกระจายคำตอบของแต่ละข้อคำถามเพียงวิธีเดียว

1.2.7 การประเมินความตรงตามโครงสร้าง จากบทความวิจัยและวิทยานิพนธ์ ที่ ทบทวนทั้งหมด 13 เรื่อง พบเพียง 5/13 เรื่องที่รายงานการประเมินความตรงเชิงโครงสร้างโดย รายละเอียดของการประเมิน ได้แก่ ความตรงเชิงผู้เข้าโดยใช้วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของ เพียร์สัน, ความตรงเชิงจำแนก ใช้วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สถิติ Mann-Whitney U test และ Independent two sample t-test การทดสอบในกลุ่มที่ทราบความแตกต่าง โดยใช้วิธี สถิติ Mann-Whitney U test หรือ Independent two sample t-test และการวิเคราะห์ปัจจัยเชิงยืนยัน (confirmatory factor analysis) ใน 5 เรื่องที่รายงานมีรายละเอียดดังนี้ 1 เรื่องใช้ทุกวิธีการยกเว้น วิธีการทดสอบในกลุ่มที่ทราบความแตกต่าง มี 1 เรื่องใช้ทุกวิธีการยกเว้นวิธีการทดสอบในกลุ่มที่ ทราบความแตกต่าง และการวิเคราะห์ปัจจัยเชิงยืนยัน มี 2 เรื่องใช้วิธีการประเมินความตรงเชิง จำแนก และมี 1 งานใช้วิธีการทดสอบในกลุ่มที่ทราบความแตกต่างโดยใช้ Mann-Whitney U test สำหรับการวิเคราะห์ปัจจัยเชิงยืนยัน พบการรายงานเฉพาะในวิทยานิพนธ์เท่านั้น

การวิเคราะห์ปัจจัยในส่วนของการประเมินความตรงเชิงโครงสร้างที่พิจารณาในส่วนนี้ จะ ใช้การวิเคราะห์ปัจจัยเชิงยืนยัน อย่างไรก็ตามในขั้นตอนของการวิเคราะห์รายการคำถาม มีการใช้ การวิเคราะห์ปัจจัยเชิงสำรวจเพื่อพิจารณาโครงสร้างของตัวแปรด้วย ดังนำเสนอแล้วในหัวข้อ 1.2.6

1.2.8 การประเมินความตรงตามเกณฑ์สัมพัทธ์ เมื่อพิจารณาจากบทความวิจัยและ วิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่ามีการรายงานประเด็นการประเมินความตรงตาม เกณฑ์สัมพัทธ์ 2/13 เรื่อง ใน 2 เรื่องนี้รายงานการประเมินความตรงร่วมสมัย (concurrent validity) โดยมีเครื่องมือมาตรฐานในการเปรียบเทียบทั้ง 2 เรื่อง สถิติที่ใช้คือสัมประสิทธิ์สหสัมพันธ์ของ เพียร์สัน

เรื่องแรกมีวิธีการประเมินความตรงร่วมสมัยโดยใช้วิธีการคำนวณค่าความสัมพันธ์ ระหว่างเครื่องมือมาตรฐานซึ่งใช้เครื่องมือ WHOQOL-BREF เทียบกับสเกลที่สร้างขึ้นคือ Pictorial Thai Quality of Life (PTQL) โดยใช้สถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สำหรับเรื่องที่สอง โดยใช้วิธีการคำนวณค่าความสัมพันธ์ระหว่างเครื่องมือมาตรฐานใช้ Health Related Self Report Scale (HRSR) เทียบกับสเกลที่สร้างขึ้น โดยใช้สถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเช่นกันซึ่ง พบการรายงานเฉพาะในบทความวิจัยเท่านั้น สำหรับงานที่เหลือไม่ได้ประเมินความตรงตามเกณฑ์ สัมพัทธ์

1.2.9 การประเมินค่าความเที่ยง จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า ทุกงานมีการรายงานค่าความเที่ยง ด้วยสัมประสิทธิ์แอลฟาของครอนบาค ซึ่งเมื่อ

จำแนกเป็นบทความและวิทยานิพนธ์ พบการรายงานในทำนองเดียวกัน ได้แก่ บทความวิจัย 3 เรื่อง และวิทยานิพนธ์ 10 เรื่อง พบว่าทุกงานมีการรายงานค่าความเที่ยง ด้วยสัมประสิทธิ์แอลฟาของครอนบาค

1.2.10 การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง (sensitivity to change) จากบทความวิจัยและวิทยานิพนธ์ ที่พบทวนทั้งหมด 13 เรื่อง ไม่พบการรายงานในประเด็นดังกล่าว

1.2.11 การรายงานการนำไปใช้ จากบทความวิจัยและวิทยานิพนธ์ ที่พบทวนทั้งหมด 13 เรื่อง พบ 6/13 เรื่อง รายงานการนำไปใช้ สิ่งทีระบุคือวิธีคิดคะแนนและการแปลผล โดยพบเฉพาะในวิทยานิพนธ์เท่านั้น รายละเอียดดังตารางที่ 3

ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์ที่ศึกษา

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความ วิจัย	วิทยานิพนธ์ระดับ บัณฑิตศึกษา	รวม
	(n=3) จำนวน	(n=10) จำนวน	(n=13) จำนวน
<b>การนิยามตัวแปรหรือกำหนดโครงสร้างเชิงเนื้อหาของสเกล</b>			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้, นำบางส่วนของสเกล เดิมที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต	-	1	1
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้	2	7	9
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, สัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการ, สนทนากลุ่มจากกลุ่มเป้าหมาย	1	1	2
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง,นำ บางส่วนของสเกลเดิมที่เกี่ยวข้องกับสเกลที่ต้องการสร้าง เพื่อวัดคุณภาพชีวิต	-	1	1

ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์  
ที่ศึกษา (ต่อ)

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความ วิจัย	วิทยานิพนธ์ระดับ บัณฑิตศึกษา	รวม
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
<b>การรวบรวมข้อคำถาม</b>			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
-เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้, และ/หรือสัมภาษณ์เชิง ลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่ม จากกลุ่มเป้าหมาย	-	2	2
-เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้	2	6	8
-เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, และ/หรือสัมภาษณ์เชิงลึก โดยระดมสมองจากผู้เชี่ยวชาญ หรือการสนทนากลุ่มจากกลุ่มเป้าหมาย	1	2	3
<b>จำนวนกลุ่มตัวอย่าง/ผู้ให้ข้อมูล (กลุ่ม)</b>			
-1 กลุ่ม	3	8	11
<b>ขนาดตัวอย่าง/จำนวนผู้ให้ข้อมูล (ค่าต่ำสุด – ค่าสูงสุด)</b>	30-160	10-28	10-160
<b>-วิธีการเก็บข้อมูล</b>			
-ให้กลุ่มเป้าหมายตอบเอง	1	-	1
-ให้กลุ่มเป้าหมายตอบเองและสัมภาษณ์โดยตรง	-	1	1
-สัมภาษณ์ความรู้สึกร่วมใจของผู้ป่วยโดยตรง	-	1	1
-สัมภาษณ์แบบเจาะลึก	-	2	2
-รวบรวมจากเครื่องมือที่มีอยู่เดิมแล้วเลือกคำถามบางข้อ	-	1	1
<b>-จำนวนคำถามเริ่มต้น (ค่าต่ำสุด – ค่าสูงสุด)</b>	18-30	22-187	18-187
<b>-จำนวนคำถามเมื่อสิ้นสุด (ค่าต่ำสุด – ค่าสูงสุด)</b>	18-22	35-83	18-83
<b>-จำนวนองค์ประกอบ (ค่าต่ำสุด – ค่าสูงสุด)</b>	4-6	4-7	4-7



ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์  
ที่ศึกษา (ต่อ)

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความ	วิทยานิพนธ์ระดับ	รวม
	วิจัย	บัณฑิตศึกษา	
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
<b>การกำหนดประเภทและรูปแบบสเกล</b>			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
-ลิเคิร์ต สเกล (Likert scale)	3	10	13
-คำตอบ 3 ระดับ	-	2	2
-คำตอบ 5 ระดับ	3	8	11
<b>การประเมินความตรงตามเนื้อหา</b>			
ไม่รายงาน	2	-	2
รายงาน	1	10	11
-ตรวจสอบจากผู้เชี่ยวชาญ	1	7	8
-ดัชนีความตรงเชิงเนื้อหา (CVI)	-	3	3
-นำไปทดสอบในกลุ่มเป้าหมาย	-	3	3
<b>การทดสอบสเกลเบื้องต้น</b>			
ไม่รายงาน	1	-	1
รายงาน	2	10	12
<b>-วิธีการทดสอบ</b>			
-สัมภาษณ์	1	4	5
-ตอบด้วยตนเอง	-	3	3
<b>-ลักษณะกลุ่มตัวอย่างในการทดสอบ</b>			
-กลุ่มตัวอย่างที่คล้ายกับประชากรเป้าหมาย	2	10	12
-ขนาดตัวอย่างในการทดสอบ (ค่าต่ำสุด – ค่าสูงสุด)	20-25	6-48	6-48

ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์  
ที่ศึกษา (ต่อ)

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความวิจัย	วิทยานิพนธ์ระดับ บัณฑิตศึกษา	รวม
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
<b>การวิเคราะห์รายการคำถาม</b>			
ไม่รายงาน	-	4	4
รายงาน	3	6	9
-การกระจายคำตอบของตัวเลือก (distribution of response)	-	5	5
-ความสัมพันธ์ระหว่างข้อคำถาม (inter-item correlation)	-	3	3
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	3	3
-ความสัมพันธ์ระหว่างข้อคำถาม กับสเกล (item-total correlation)	-	2	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	2	2
-ความสอดคล้องภายใน (internal consistency)	1	4	5
-รายงานค่าสัมประสิทธิ์แอลฟาของครอนบาค			
-ภาพรวม	1	4	5
-รายค้าน	1	1	2
-การวิเคราะห์ปัจจัย (factor analysis)	3	5	8
-การวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis)	3	5	8
<b>การประเมินความตรงตามโครงสร้าง</b>			
ไม่รายงาน	1	7	8
รายงาน	2	3	5
-ความตรงเชิงผู้เข้า	-	2	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	2	2
-ความตรงเชิงจำแนก	-	4	4
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	2	2	4

ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์  
ที่ศึกษา (ต่อ)

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความ วิจัย (n=3)	วิทยานิพนธ์ระดับ บัณฑิตศึกษา (n=10)	รวม (n=13)
	จำนวน	จำนวน	จำนวน
-การวิเคราะห์ปัจจัย	-	1	1
-การวิเคราะห์ปัจจัยเชิงยืนยัน (confirmatory factor analysis)	-	1	1
-การทดสอบในกลุ่มที่ทราบความแตกต่าง	-	1	1
-วิธีสถิติ Mann-Whitney U test	-	1	1
การประเมินความตรงตามเกณฑ์สัมพัทธ์			
ไม่รายงาน	1	10	11
รายงาน	2	-	2
-ความตรงร่วมสมัย	2	-	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	2	-	-
รายละเอียดรายงานความตรงร่วมสมัย			
-ไม่มี	-	-	-
- มี เช่น PTQL (เครื่องมือ) และ WHOQOL, เครื่องมือที่สร้างขึ้นกับ Health related self report scale (HRSR)	2	-	2
-ความตรงเชิงทำนาย	-	-	-
-วิธีสถิติ ใช้การวิเคราะห์ถดถอย	-	-	-
การประเมินความเที่ยง			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
-ความสอดคล้องภายใน	3	10	13
-วิธีการแบ่งครึ่ง	-	-	-
-วิธีสัมประสิทธิ์แอลฟา	3	10	13
รายงานค่าสัมประสิทธิ์แอลฟา			
-ภาพรวม	2	9	11
-รายด้าน	3	6	9

ตารางที่ 3 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์  
ที่ศึกษา (ต่อ)

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความ วิจัย (n=3)	วิทยานิพนธ์ระดับ บัณฑิตศึกษา (n=10)	รวม (n=13)
	จำนวน	จำนวน	จำนวน
-ความคงที่ของการวัด	-	1	1
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	1	1
การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง			
ไม่รายงาน	3	10	13
รายงาน	-	-	-
การรายงานการนำไปใช้			
ไม่รายงาน	3	4	7
รายงาน	-	6	6
-คู่มือการใช้	-	-	-
-วิธีการคิดคะแนนและการแปลผล	-	6	6
-รายละเอียดของหน่วยงานหรือบุคคลที่ใช้ติดต่อ เพื่อขอเช่าเครื่องมือไปใช้	-	-	-

## 2. การอภิปรายผล

จากการศึกษาการรายงานระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ จากบทความวิจัยในวารสารด้านการแพทย์และสาธารณสุขของไทย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ทางกายภาพและสาธารณสุขของมหาวิทยาลัยในประเทศไทย ที่พิมพ์เผยแพร่ระหว่างเดือนมกราคม พ.ศ.2543 ถึงเดือนธันวาคม พ.ศ.2552 โดยมีประเด็นในการพิจารณา 11 ประเด็น ซึ่งผู้วิจัยได้ศึกษาวิธีการสร้างสเกลจากตำราหลายเล่ม ได้แก่ Spector (1992) Devellis (2003) Streiner and Norman (2003) และ Fayers and Machin (2000) และบทความวิจัยประเภทบททวนระเบียบวิธีการสร้างเครื่องมือประเภทสเกลเพื่อวัดตัวแปรเชิงโครงสร้าง (construct) (Hinkin, 1995; Hensley, 1999; Costes, 1995; Costes, 1997) ซึ่งจะนำเสนอการอภิปรายดังนี้

## 2.1 ลักษณะทั่วไปของบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษา

บทความวิจัยและวิทยานิพนธ์ด้านการแพทย์และสาธารณสุขที่นำมาศึกษา มีทั้งหมด 13 เรื่อง พบว่า 9/13 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อสร้างสเกลวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ 4/13 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อประเมินคุณภาพชีวิตแต่ผู้วิจัยสร้างสเกลวัดคุณภาพชีวิตขึ้นใหม่ ซึ่งพบว่า 9/13 เรื่อง สร้างสเกลเพื่อวัดตัววัดสุขภาพเกี่ยวกับโรคเรื้อรัง เช่น คุณภาพชีวิตผู้มีปัญหาสายตา คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด ที่เหลือเป็นตัววัดสุขภาพในประชากรเฉพาะกลุ่ม ได้แก่ คุณภาพชีวิตสตรีวัยหมดประจำเดือน ผู้เลิกบุหรี่ และผู้ป่วยเด็ก

วัตถุประสงค์ของการศึกษาเพื่อสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพพบว่าตัววัดด้านสุขภาพที่ต้องการสร้าง ได้แก่ คุณภาพชีวิตผู้มีปัญหาด้านสายตา คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด เป็นต้น ซึ่ง จิราพร เขียวอยู่ (2552) ได้สรุปว่าตัววัดด้านสุขภาพที่ใช้ในงานสาธารณสุข เพื่อบ่งบอกปัญหาสุขภาพของประชากรใช้วางแผนงานด้าน โยบายสุขภาพและวางแผนการรักษาทางการแพทย์ พบว่า หลายตัววัดไม่สามารถวัดได้โดยตรงจากคำถามเพียงข้อเดียว ตัววัดลักษณะนี้เรียกว่าตัววัดที่มีโครงสร้างจำเป็นต้องใช้คำถามหลายข้อประกอบกัน แล้วนำคำตอบจากคำถามทุกข้อรวมกันจึงจะได้ค่าที่บอกลักษณะของบุคคลที่ถูกวัด

## 2.2 การรายงานระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างสเกลตามประเด็นที่พิจารณา

ผลจากการพิจารณาระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้ในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพตามมาตรฐานการสร้างสเกลที่ปรากฏในงานวิจัยที่ทบทวนมีข้อสรุปและการอภิปรายผลดังนี้

### 2.2.1 การกำหนดแนวคิด หรือ โครงสร้างของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด

เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง ทุกงานรายงานว่าทำขั้นตอนนี้ โดยมี 1 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต มี 9 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา มี 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 1 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต จะเห็นวิธีการที่ปรากฏในงานที่ทบทวน สอดคล้องกับคำแนะนำ

ของ Streiner and Norman (2003) และ Spector (1992) แนะนำไว้ว่า ในการกำหนดโครงสร้างของตัวแปรเพื่อสร้างสเกลเพื่อวัดนั้น ควรใช้หลายวิธีประกอบกัน ได้แก่ การทบทวนแนวคิดทฤษฎีที่เกี่ยวข้อง การศึกษาจากงานวิจัยที่ผ่านมา การนำบางส่วนของงานในทำนองเดียวกันมาใช้ ตลอดจนการสอบถามความเห็นจากผู้เชี่ยวชาญและกลุ่มเป้าหมายด้วยวิธีการสัมภาษณ์เชิงลึกและหรือการสนทนากลุ่ม และเป็นไปในทำนองเดียวกับการศึกษาของ Hinkin (1995) ซึ่งทบทวนบทความ 75 บทความ พบ ร้อยละ 83 (62/75) ของบทความที่ทบทวนมีการกำหนดนิยามของสิ่งที่ต้องการวัดเพื่อนำมากำหนดโครงสร้างเชิงเนื้อหาของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด โดยใช้วิธีการทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา ร้อยละ 11 (8/75) ของบทความที่ทบทวนใช้วิธีการสนทนากลุ่มจากกลุ่มเป้าหมาย และ Hensley (1999) ทบทวนบทความ 6 บทความ พบว่า 6/6 ของบทความที่ทบทวน รายงานการกำหนดโครงสร้างเชิงเนื้อหาของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด โดยใช้วิธีการทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา อย่างไรก็ตาม ในงานที่ทบทวนนี้มีสิ่งที่พบแตกต่างกัน คือ บางงานใช้หลายวิธีร่วมกัน แต่บางงานใช้เพียงหนึ่งหรือสองวิธีข้างต้น

### 2.2.2 การรวบรวมรายการคำถาม

เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง ทุกงานรายงานว่าทำขั้นตอนนี้ โดยมี 2 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และ/หรือสัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 3 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย วิธีการที่การศึกษา 5 การศึกษานี้ใช้ได้รับการยอมรับว่าเป็นวิธีการที่เหมาะสมที่สุด (Streiner and Norman, 2003) และมี 8 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา วิธีการที่ 8 งานนี้ใช้เป็นวิธีที่นักวิชาการด้านการสร้างสเกลหลายคนแนะนำ (Spector, 1992; Devellis, 2003; Streiner and Norman, 2003; Fayers and Machin, 2000) ดังนั้นในประเด็นของการรวบรวมรายการคำถามนี้ จึงสรุปได้ว่า งานวิจัยที่ทบทวนใช้วิธีการที่หลากหลายประกอบกัน ซึ่งถือว่าสอดคล้องกับหลักการสร้างเครื่องมือ นั่นคือในขั้นต้นผู้สร้างกำหนดรายการคำถาม (item pool) ให้มีปริมาณมากที่สุด เพื่อครอบคลุมเนื้อหาของสิ่งที่ต้องการวัด และเพิ่มความเที่ยงของเครื่องมือ (Spector, 1992; Devellis, 2003; Streiner and Norman, 2003; Fayers and Machin, 2000)

### 2.2.3 การกำหนดประเภทและรูปแบบของสเกล

เมื่อพิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า ทั้งหมดใช้สเกลของลิเคิร์ต (Likert scale) โดย 11/13 เรื่อง กำหนดตัวเลือก

ของคำตอบเป็น 5 ระดับ ซึ่งสอดคล้องกับ Hinkin (1995) ที่พบทวนบทความ 75 บทความ แล้วพบ 36/75 ของบทความที่ทบทวน กำหนดประเภทสเกลและรูปแบบคำตอบเป็นเช่นนี้ ส่วนงานของ Hensley (1999) พบว่า 5/6 ของบทความที่ทบทวน ใช้สเกลของลิเคิร์ตเช่นกัน แต่กำหนดคำตอบเป็น 5-7 ระดับ ส่วนบทความวิจัยและวิทยานิพนธ์ที่เหลืออีก 2 งาน ใช้สเกลของลิเคิร์ตเช่นกัน แต่กำหนดคำตอบเป็น 3 ระดับ โดยพบในการสร้างสเกลเพื่อวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพในกลุ่มผู้ป่วยเด็กและผู้ป่วยหลังผ่าตัดเปิดกะโหลกศีรษะ

สเกลของลิเคิร์ตมีลักษณะเฉพาะคือ ประกอบด้วยรายการคำถามหลายข้อ แต่ละข้อวัดออกมาเป็นเชิงปริมาณได้ ไม่มีคำตอบที่ถูกหรือผิดแต่จะให้ผู้ตอบประเมินค่า และแต่ละรายการคำถามจะมีความสัมพันธ์กันเนื่องจากทุกรายการคำถามมุ่งวัดสิ่งเดียวกัน ดังนั้นการเลือกใช้สเกลประเภทนี้ จึงมีส่วนช่วยให้เครื่องมือที่สร้างขึ้นมีความตรง (validity) และความเที่ยง (reliability) สูง (จิราพร เขียวอยู่, 2552)

การกำหนดระดับของตัวเลือกของคำตอบมีตั้งแต่ 3-7 ระดับ ขึ้นอยู่กับผู้สร้าง ที่พบบ่อยคือ 4-5 ระดับ (Fayers and Machin, 2000) การกำหนดตัวเลือกของคำตอบจำนวนน้อย เช่น 3 ระดับ เครื่องมือจะมีความเที่ยงต่ำกว่ากรณีกำหนดตัวเลือกของคำตอบจำนวนมากเช่น 5 ระดับ ดังนั้นควรกำหนดประมาณ 5-9 ระดับ (Spector, 1992)

การกำหนดรูปแบบสเกลเป็นสเกลของลิเคิร์ตนี้ มักเป็นทางเลือกในการสร้างเครื่องมือ ที่มีวัตถุประสงค์เพื่อการประเมิน (evaluative instrument) หรือเพื่อใช้วัดการเปลี่ยนแปลงของบุคคลหรือกลุ่มตัวอย่างในช่วงเวลาหนึ่ง (Fayers and Machin, 2000)

ดังนั้น จึงกล่าวสรุปได้ว่า การกำหนดประเภทและรูปแบบสเกลของงานที่ทบทวนทั้งหมดเหมาะสม ซึ่งถึงแม้จะมี 2 งานที่กำหนดตัวเลือกของคำตอบเป็น 3 ระดับ ก็ไม่จัดว่าไม่เหมาะสม เพราะใน 2 การศึกษานั้น งานหนึ่งศึกษาในผู้ป่วยหลังผ่าตัดเปิดกะโหลกศีรษะ อีกงานหนึ่งศึกษาในผู้ป่วยเด็ก ซึ่งกลุ่มเป้าหมายของทั้งสองงานนี้ หากใช้ตัวเลือกของคำตอบมากกว่า 3 ระดับ ผู้ตอบก็อาจไม่สามารถแยกแยะคำตอบได้

#### 2.2.4 การประเมินความตรงตามเนื้อหา

การประเมินความตรงตามเนื้อหา (content validity) เป็นการพิจารณาว่าเครื่องมือมีเนื้อหาครอบคลุมทุกประเด็นของคุณสมบัติของตัวแปรที่ต้องการวัดหรือไม่ โดยเป็นเนื้อหาที่สอดคล้องกับแนวคิดและเป็นไปตามคำนิยามเชิงปฏิบัติการของตัวแปรนั้น จากการศึกษาครั้งนี้พบว่า มี 2/13 เรื่อง ไม่รายงานการประเมินความตรงด้านเนื้อหา ที่เหลือ 11/13 รายงานว่าประเมินความตรงด้านเนื้อหา แต่รายละเอียดของวิธีการที่ใช้แตกต่างกัน โดยพบดังนี้ 3/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยวิธีการนำไปทดสอบในกลุ่มเป้าหมาย ทำได้โดยนำเครื่องมือที่

สร้างขึ้น ไปสัมภาษณ์กับกลุ่มเป้าหมายโดยตรง พิจารณาความเข้าใจในการตอบข้อคำถาม หากกลุ่มเป้าหมายไม่เข้าใจข้อความใดจะตัดข้อคำถามนั้นออกไป และการเลือกคำตอบจากตัวเลือกในแต่ละข้อคำถามถ้ามีการเลือกคำตอบในข้อใดข้อหนึ่งมากเกินไปเกินร้อยละ 90 จะตัดข้อคำถามข้อนั้นออกไป 8/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยใช้วิธีขอความเห็นจากผู้เชี่ยวชาญ กรณีเน้นวิธีการเชิงคุณภาพ คือให้ผู้เชี่ยวชาญให้ความเห็นในส่วนต่าง ๆ ของเครื่องมือ แล้วผู้วิจัยนำความเห็นของผู้เชี่ยวชาญมาปรับแก้เครื่องมือ การประเมินที่พบในลักษณะนี้ สอดคล้องกับผลการศึกษาของ Coste (1997) ที่พบทวนบทความจำนวน 42 บทความ พบว่า 2/42 ของบทความที่ทบทวน รายงานการประเมินความตรงตามเนื้อหาด้วยวิธีขอคำแนะนำจากผู้เชี่ยวชาญ ส่วนงานของ Hensley (1999) และ Coste (1995) ก็พบว่าใช้วิธีการตรวจสอบเชิงคุณภาพโดยผู้เชี่ยวชาญเช่นกัน, และมีงานที่ทบทวน 3/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยใช้วิธีขอความเห็นจากผู้เชี่ยวชาญร่วมกับการคำนวณค่าดัชนีความตรงด้านเนื้อหา (content validity index: CVI) แล้วสรุปว่าค่า CVI เป็นไปตามเกณฑ์ของ Lynn นั่นคือ เมื่อให้ผู้เชี่ยวชาญให้ระดับความเห็นด้านความตรงเชิงเนื้อหาเป็น 4 ระดับ ควรมีสัดส่วนผู้เชี่ยวชาญที่เห็นด้วยกับข้อคำถาม โดยให้ระดับการประเมินระดับที่ 3 หรือ 4 อย่างน้อยร้อยละ 80 ของผู้เชี่ยวชาญทั้งหมด Lynn (1986, อ้างใน Streiner and Norman, 2003)

สำหรับจำนวนผู้เชี่ยวชาญที่ใช้ในการตรวจสอบความตรงด้านเนื้อหา พบว่า มีจำนวนระหว่าง 3-9 คน โดยส่วนใหญ่ใช้ผู้เชี่ยวชาญ 5 คน

แม้จะพบว่างานที่ทบทวนส่วนใหญ่ มีการประเมินความตรงตามเนื้อหา แต่ก็มี 2 บทความวิจัยที่ไม่มีการรายงานในส่วนนี้ ซึ่งผู้ศึกษาไม่แน่ใจว่า การไม่ได้รายงาน เป็นเพราะไม่ได้ทำ หรือทำแต่ไม่ได้เขียนรายละเอียดลงในบทความวิจัย อย่างไรก็ตาม หากยึดตามสิ่งที่ปรากฏในบทความวิจัย ก็ถือว่าส่วนหนึ่งของงานที่ทบทวนยังไม่ได้ใช้ระเบียบวิธีมาตรฐานในการสร้างเครื่องมือ เพราะขั้นตอนนี้เป็นสิ่งที่จำเป็นต้องทำ

#### 2.2.5 การทดสอบสเกลเบื้องต้น

การทดสอบสเกลเบื้องต้นมีเป้าหมายเพื่อตรวจสอบความเข้าใจคำถามของกลุ่มเป้าหมาย (Fayers and Machin, 2000) และนำข้อมูลที่ได้ไปวิเคราะห์รายการคำถาม จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบ 12/13 เรื่อง รายงานการทดสอบสเกลเบื้องต้นจำนวน โดยนำไปทดสอบกับกลุ่มคนที่มีลักษณะเหมือนประชากรเป้าหมาย จาก 12 เรื่องของงานที่ทบทวนในการวิจัยครั้งนี้ มี 5/12 เรื่องใช้การสัมภาษณ์จากตัวอย่างที่มีลักษณะเหมือนประชากรเป้าหมาย ขนาดตัวอย่างที่ใช้มีจำนวน 30, 47, 52 และ 60 คน ตามลำดับ และ 3/12 เรื่อง

กลุ่มตัวอย่างตอบด้วยตนเอง ขนาดตัวอย่างที่ใช้มีจำนวน 6, 20, 30 คน ตามลำดับ ส่วนงานที่เหลือไม่ระบุวิธีการ

การทดสอบสเกลขั้นต้นที่มีเป้าหมายเพื่อตรวจสอบความเข้าใจภาษาในรายการคำถามของกลุ่มเป้าหมาย ต้องการขนาดตัวอย่างจำนวนไม่มาก คือประมาณ 10 – 30 คน (Fayers and Machin, 2000) พิจารณาเฉพาะส่วนนี้สรุปได้ว่างานเกือบส่วนใหญ่ (8/12 เรื่อง) ที่พบทวนในการศึกษานี้ ใช้วิธีการที่เหมาะสม

### 2.2.6 การวิเคราะห์รายการคำถาม

การวิเคราะห์รายการคำถามมีเป้าหมายเพื่อ ลดข้อคำถามให้เหลือเฉพาะที่จำเป็นเท่านั้น คือ ครอบคลุมเนื้อหา และไม่มากเกินไปจนต้องใช้เวลาในการตอบคำถาม การวิเคราะห์รายการคำถาม ทำให้หลายวิธีต่อไปนี้ (จิราพร เขียวอยู่, 2548) 1) การพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม (distribution of response) 2) วิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม (inter-item correlation) 3) วิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม (item-total correlation) 4) วิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถาม (inter-item consistency) และ 5) การวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis) เพื่อหาโครงสร้างของข้อมูล โดยวิธีที่ 2) และ 3) นั้นอาจจะไม่ดำเนินการก็ได้ ข้อสรุปเกี่ยวกับวิธีการวิเคราะห์รายการคำถามข้างต้น เป็นไปตามข้อเสนอแนะจากเอกสารตำราหลายรายการ ได้แก่ Streiner and Norman (2003) Spector (1992) DeVellis (2003) และ Nunnally (1978)

ในการศึกษาคั้งนี้ เมื่อพิจารณาบทความวิจัยและวิทยานิพนธ์ ที่พบทวนทั้งหมด 13 เรื่อง พบว่า 4/13 เรื่อง ไม่รายงาน ที่เหลือ 9/13 เรื่องรายงาน ใน 9 เรื่องที่รายงาน มีรายละเอียดคั้งนี้ 1 เรื่อง วิเคราะห์รายการคำถามทุกวิธียกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม, มี 2 เรื่อง ใช้วิธีวิเคราะห์รายการคำถามทุกวิธีคั้งเสนอข้างต้น โดยสถิติที่ประเมินความสัมพันธ์คือสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และสถิติที่ใช้ประเมินความสอดคล้องภายในคือสัมประสิทธิ์แอลฟาของครอนบาค 1 เรื่อง ใช้วิธีวิเคราะห์รายการคำถามทุกวิธีคั้งเสนอข้างต้น ยกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม 1 เรื่อง ใช้วิธีการพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม และการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล 1 เรื่อง ใช้วิธีการวิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถามและการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล 2 เรื่อง ใช้วิธีการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล 1 เรื่อง ไม่รายงานวิธีการที่ใช้

จะเห็นว่ากรณีการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล พบ 8/9 เรื่องที่ใช้วิธีการนี้ มีรายละเอียดของการวิเคราะห์คั้งนี้ ทุกงานใช้การสกัดปัจจัยด้วยวิธีองค์ประกอบ

หลัก (principle component analysis) มีการคัดเลือกจำนวนปัจจัยที่เหมาะสมโดยพิจารณาจากค่า eigenvalue ที่มากกว่า 1 ในจำนวนนี้เลือกวิธีการหมุนแกนแบบมุมแหลม (oblique) ด้วยวิธี promax จำนวน 2 เรื่อง ที่เหลือจำนวน 6 เรื่องใช้วิธีการหมุนแกนแบบมุมฉาก (orthogonal) โดยใช้วิธี varimax 5 เรื่องและวิธี quatimax 1 เรื่อง สำหรับเกณฑ์ในการคัดเลือกข้อคำถามในแต่ละปัจจัยจะมีเกณฑ์พิจารณาจากค่า factor loading ที่มีค่ามากกว่า 0.4 ข้อคำถามจะคงอยู่ในปัจจัยนั้น พบ 7 งานที่ใช้เกณฑ์นี้ และที่เหลือ 1 งาน ใช้เกณฑ์จากค่า factor loading ที่มีค่า 0.7 ขึ้นไปโดยขนาดตัวอย่างที่ใช้ในขั้นตอนนี้มีจำนวนระหว่าง 47 ถึง 5,574 คน

ในงานที่ทบทวนนี้ได้รายงานการกำหนดขนาดตัวอย่างไว้มีจำนวนระหว่าง 47 ถึง 5,574 คน (47, 50, 73, 100, 107, 111, 174, 182, 252, 424, 500, 672 และ 5,574) ซึ่งในการวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis; EFA) เพื่อหาโครงสร้างของข้อมูล จำเป็นต้องใช้ขนาดตัวอย่างจำนวนมาก จึงควรกำหนดขนาดตัวอย่างให้เพียงพอต่อการวิเคราะห์ที่ Comrey (1973) เสนอหลักการทั่วไปในการกำหนดขนาดตัวอย่างไว้ดังนี้ ขนาดตัวอย่าง 100 คน ถือว่าน้อย 200 คนถือว่าพอใช้ 300 คน ถือว่า ดี 1,000 คน ถือว่ายอดเยี่ยม Tinsley และ Tinsley (1987) และ Hatcher (1994) เสนอขนาดตัวอย่างต่ำสุดที่เพียงพอในการวิเคราะห์คือ 100 คนขึ้นไป หรือคิดเป็น 5-10 เท่าของจำนวนข้อคำถาม Guadagnoli และ Velicer (1988) ได้เสนอการกำหนดขนาดตัวอย่างสำหรับการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูลจากขนาดของน้ำหนักองค์ประกอบ (factor loading) และจำนวนข้อคำถามที่มีอยู่ต่อหนึ่งด้านดังนี้ ถ้าค่าน้ำหนักองค์ประกอบมีค่า 0.8 ขึ้นไปไม่จำเป็นต้องคำนึงถึงจำนวนข้อคำถามที่มีอยู่ในหนึ่งด้าน นั่นคือจำนวนขนาดตัวอย่างที่ต้องการจะน้อยกว่าหรือเท่ากับ 50 คนก็เพียงพอ ถ้าค่าน้ำหนักองค์ประกอบมีค่าอยู่ในช่วง 0.60 ต้องใช้ขนาดตัวอย่างประมาณ 150 คนขึ้นไป ถ้าค่าน้ำหนักองค์ประกอบมีค่าในช่วง 0.40 และมีจำนวนข้อคำถามตั้งแต่ 10 ข้อขึ้นไปในแต่ละด้านต้องใช้ขนาดตัวอย่างประมาณ 150 คนขึ้นไป ถ้ามีจำนวนข้อคำถามน้อยกว่า 10 ข้อ แต่ละด้านต้องใช้ขนาดตัวอย่าง 300 – 400 คน สำหรับผลจากการทบทวนงานวิจัยประเภททบทวนระเบียบวิธีการสร้างสเกลชนิดใหม่ทั้ง 13 งานนี้ มีการกำหนดขนาดตัวอย่างในการวิเคราะห์ต่ำสุดมี 3 งาน (บุขจรี วุฒิสุพงษ์, 2541; เตือนใจ เห่งจิ้น, 2542; เพ็ญพิมล ยิงยง, 2550) จำนวน 47, 50 และ 73 คนตามลำดับ เมื่อพิจารณาการกำหนดขนาดตัวอย่างตามเกณฑ์ของ Comrey (1973) ถือว่ายังน้อยเกินไป

สำหรับการวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล วิธีการนี้จะทำการวิเคราะห์จัดกลุ่มข้อคำถามและเรียกกลุ่มเหล่านั้นว่า องค์ประกอบหรือด้านหรือปัจจัยหรือมิติ ทำให้เห็นโครงสร้างของตัวแปรที่ศึกษา โดยข้อคำถามที่อยู่ในองค์ประกอบเดียวกันจะมี

ความสัมพันธ์กันมาก ส่วนข้อคำถามที่อยู่ต่างองค์ประกอบกันจะมีความสัมพันธ์กันน้อยหรือไม่มี  
ความสัมพันธ์กัน

ขั้นตอนแรกของการวิเคราะห์ปัจจัยเพื่อหาโครงสร้างของข้อมูลคือ การพิจารณา  
ความเหมาะสมของข้อมูลตัวอย่างที่นำมาวิเคราะห์ปัจจัยเพื่อหาโครงสร้างของข้อมูล โดยพิจารณา  
ว่าข้อคำถามมีความสัมพันธ์กันหรือไม่ เนื่องจากในการวิเคราะห์ปัจจัยเพื่อหาโครงสร้างของข้อมูล  
จะได้ผลดีเมื่อตัวแปรที่นำมาวิเคราะห์มีความสัมพันธ์กันสูง ถ้าสัมพันธ์กันน้อยอาจไม่มี  
องค์ประกอบร่วมกัน โดยค่าสหสัมพันธ์ระหว่างข้อคำถามควรมีค่ามากกว่าหรือเท่ากับ 0.3  
(Tabachnick และ Fidell, 2007) สำหรับผลจากการทบทวนงานวิจัยประเภททบทวนระเบียบวิธีการ  
สร้างสเกลชนิดใหม่ทั้ง 13 งานนี้ มีการรายงานการพิจารณาความเหมาะสมของข้อมูลตัวอย่างจาก  
ค่าสหสัมพันธ์เท่ากับ 0.3 จำนวน 6 งาน (ฐิติพร นาคทวน, 2546; สิริรัตน์ โพธิ์ศิริสุข, 2545; จารุณี  
เจริญแสงรุ่ง, 2542; เพ็ญญา อุ่ณสนธิ, 2542; เพ็ญพิมล ยิ่งยง, 2550; สุชีรา ภัทรายุตวรรตน์, 2548)  
เมื่อพิจารณาตามเกณฑ์ของ Tabachnick และ Fidell (2007) ถือว่ามีความเหมาะสม

ขั้นต่อมาคือ การกำหนดวิธีการสกัดปัจจัย ซึ่งเป็นการหาจำนวนปัจจัย ที่สามารถ  
ใช้เป็นตัวแทนของตัวแปรทั้งหมดที่มีอยู่ โดยปัจจัยนั้นต้องอธิบายความแปรผันร่วมระหว่างตัวแปร  
(ข้อคำถาม) ได้มากที่สุด จากการทบทวนงานวิจัยประเภททบทวนระเบียบวิธีการสร้างสเกลชนิด  
ใหม่ทั้ง 13 งานนี้พบว่าทั้ง 8 งาน (ฐิติพร นาคทวน, 2546; สิริรัตน์ โพธิ์ศิริสุข, 2545; จารุณี  
เจริญแสงรุ่ง, 2542; เพ็ญญา อุ่ณสนธิ, 2542; เพ็ญพิมล ยิ่งยง, 2550; สุชีรา ภัทรายุตวรรตน์, 2548;  
นงพรรณ มโนยศ, 2549; พยงค์ หาญผดุงกิจ, 2544) ใช้การสกัดปัจจัยด้วยวิธีองค์ประกอบหลัก  
(principle component analysis) และใน 8 งานนี้ใช้เกณฑ์การตัดสินใจเลือกจำนวนปัจจัยจากค่า  
eigenvalue มากกว่าหรือเท่ากับ 1 ซึ่งถือว่ามีความเหมาะสม (Tabachnick และ Fidell, 2007)

หลังจากกำหนดจำนวนด้านที่เหมาะสมแล้ว ต่อมาเป็นขั้นตอนการหมุนแกน  
ปัจจัย (factor rotation) เนื่องจากผลจากการสกัดปัจจัยอาจยังไม่ได้รูปแบบปัจจัยที่ชัดเจน จึงต้องทำ  
ขั้นตอนนี้ ซึ่งการหมุนแกนองค์ประกอบมี 2 แบบคือ การหมุนแบบมุมฉาก (orthogonal rotation)  
เมื่อหมุนแกนแล้วปัจจัยจะเป็นอิสระต่อกัน และการหมุนแกนแบบมุมแหลม (oblique rotation) เมื่อ  
หมุนแล้วปัจจัยจะมีความสัมพันธ์กันหรือไม่เป็นอิสระต่อกัน การเลือกใช้วิธีการใดขึ้นอยู่กับเนื้อหา  
ของตัวแปรที่ต้องการวัดนั้น จากงานที่ทบทวนทั้ง 8 งาน เลือกวิธีการหมุนแกนแบบมุมแหลม  
จำนวน 2 เรื่อง โดยหมุนแกนด้วยวิธี promax (ฐิติพร นาคทวน, 2546; เพ็ญญา อุ่ณสนธิ, 2542)  
เนื่องจากงานทั้งสองนี้ต้องการผลลัพธ์เป็นปัจจัยที่มีความสัมพันธ์กันและไม่ต้องการนำไปวิเคราะห์  
ด้วยสถิติวิธีอื่นต่อไป ส่วนที่เหลือจำนวน 6 เรื่องใช้วิธีการหมุนแกนแบบมุมฉาก (orthogonal) ด้วย  
วิธี varimax 5 เรื่อง (จารุณี เจริญแสงรุ่ง, 2542; เพ็ญพิมล ยิ่งยง, 2550; สุชีรา ภัทรายุตวรรตน์, 2548;

นงพรรณ มโนยศ, 2549; พยนต์ หาญผดุงกิจ, 2544) และวิธี quatimax 1 เรื่อง (สิริรัตน์ โพธิ์ศิริสุข, 2545) ซึ่งทั้ง 6 งานนี้ต้องการผลลัพธ์เป็นปัจจัยที่ไม่สัมพันธ์กันและนำไปเป็นตัวแปรสำหรับการวิเคราะห์ทางสถิติอื่น ๆ เช่น การวิเคราะห์ถดถอยเพื่อศึกษาคุณภาพชีวิตของคนไข้ต่อไป

นอกจากนี้พบว่า 5/9 เรื่องใช้การวิเคราะห์ปัจจัยเชิงสำรวจเพื่อหาโครงสร้างของข้อมูล ร่วมกับการวิเคราะห์ความสอดคล้องภายใน ซึ่งถือว่าเป็นการใช้ระเบียบวิธีที่เหมาะสม สถิติที่ใช้ประเมินความสอดคล้องภายใน คือ สัมประสิทธิ์แอลฟาของครอนบาค ค่าที่ได้อยู่ระหว่าง 0.74 ถึง 0.98 สำหรับการพิจารณาการกระจายของคำตอบซึ่งเป็นกระบวนการเริ่มต้นของการวิเคราะห์รายการคำถามที่ทุกงานควรทำนั้น มี 5/9 งานที่ดำเนินการและพบว่าใน 5 งานนี้ทำการวิเคราะห์ปัจจัยด้วย ดังนั้นจึงกล่าวได้ว่า งานที่ทบทวนในการศึกษารั้งนี้ ใช้วิธีที่เหมาะสมสำหรับการวิเคราะห์รายการคำถาม ผลการศึกษานี้สอดคล้องกับการศึกษาของ Hinkin (1995) ที่พบว่าวิธีการวิเคราะห์รายการคำถามที่ใช้ได้แก่ การประเมินความสอดคล้องภายในด้วยสถิติสัมประสิทธิ์แอลฟาของครอนบาค (ร้อยละ 97 (73/75) ของบทความที่ทบทวน) และการวิเคราะห์ปัจจัยเชิงสำรวจ (ร้อยละ 33 (25/75) ของบทความที่ทบทวน) และการศึกษาของ Hensley (1999) ที่พบว่า 6/6 บทความ รายงานการประเมินความสอดคล้องภายในด้วยสถิติสัมประสิทธิ์แอลฟาของครอนบาค และ 3/6 บทความ รายงานว่าใช้การวิเคราะห์ปัจจัยยืนยัน

### 2.2.7 การประเมินความตรงเชิงโครงสร้าง

จากบทความวิจัยและวิทยานิพนธ์ที่ทบทวนทั้งหมด 13 เรื่อง พบเพียง 5/13 เรื่องที่รายงานการประเมินความตรงเชิงโครงสร้างโดยรายละเอียดของการประเมิน ได้แก่ ความตรงเชิงผู้เข้าโดยใช้สถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ความตรงเชิงจำแนก ใช้สถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สถิติ Mann-Whitney U test และ Independent two sample t-test การทดสอบในกลุ่มที่ทราบความแตกต่างโดยใช้วิธีสถิติ Mann-Whitney U test หรือ Independent two sample t-test และการวิเคราะห์ปัจจัยยืนยัน (confirmatory factor analysis) ใน 5 เรื่องที่รายงานมีรายละเอียดดังนี้ 1 เรื่องใช้ทุกวิธีการยกเว้นวิธีการทดสอบในกลุ่มที่ทราบความแตกต่าง มี 1 เรื่องใช้ทุกวิธีการ ยกเว้นวิธีการทดสอบในกลุ่มที่ทราบความแตกต่างและการวิเคราะห์ปัจจัย มี 2 เรื่องใช้วิธีการประเมินความตรงเชิงจำแนก และมี 1 งานใช้วิธีการทดสอบในกลุ่มที่ทราบความแตกต่างด้วยสถิติ Mann-Whitney U test

สำหรับการวิเคราะห์ปัจจัยยืนยัน ซึ่งพบการรายงานเฉพาะในวิทยานิพนธ์ (รุ่งทิพย์ สุขวรรณ, 2551) เท่านั้น โดยผู้สร้างเครื่องมือให้รายละเอียดของการประเมินความตรงตามโครงสร้างว่า เป็นการตรวจสอบว่าในแต่ละมิติและข้อคำถามแต่ละข้อของ Refractive error quality



of life scale for Thai adults (REQ-Thai) ที่สร้างขึ้นจากแนวคิดทฤษฎีและการสนทนากลุ่ม มีความสอดคล้องกับข้อมูลจากการศึกษาหรือไม่ แต่ไม่พบรายละเอียดของวิธีการทางสถิติหรือเกณฑ์การพิจารณาใด ๆ ทั้งนี้ ผู้สร้างอาจมีเกณฑ์แต่ไม่อธิบายรายละเอียดจึงไม่พบในงานที่ทบทวน อย่างไรก็ตาม ได้มีข้อเสนอแนะจากเอกสาร ดำรง เกี่ยวกับการพิจารณาแบบจำลองที่เหมาะสม จะพิจารณาจากเกณฑ์ของค่า goodness of fit test ดังนี้ model chi-square ที่มี p-value มากกว่า 0.05, comparative fit index (CFI) มากกว่า 0.95, root mean square error of approximation (RMSEA) และ standardized root mean square error of approximation (SRMR) น้อยกว่า 0.05, Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), expected cross validation index (ECVI) มีค่าต่ำที่สุด (Byrne, 2001; Kline, 2005)

สำหรับวิธีการที่ใช้ประเมินความตรงเชิงโครงสร้างในงานที่ทบทวนครั้งนี้ สอดคล้องกับวิธีการที่พบในการศึกษาของ Hinkin (1995) Hensley (1999) Coste (1995) และ Coste (1997) โดยในงานที่ Hinkin (1995) ทบทวน พบ 7/75 เรื่อง และ 6/75 เรื่อง ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงกลุ่มเข้าและความตรงเชิงจำแนก ตามลำดับ งานที่ Hensley (1999) ทบทวน พบอย่างละ 1/6 เรื่อง เท่ากัน ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงกลุ่มเข้า และความตรงเชิงจำแนก สำหรับการศึกษาของ Coste (1995) และ Coste (1997) พบการรายงานการประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงกลุ่มเข้าในภาพรวม ดังนั้นจึงกล่าวได้ว่า ในงานที่ทบทวนครั้งนี้ในส่วนของงานที่มีการประเมินความตรงเชิงโครงสร้างนั้น ได้ใช้วิธีการที่เหมาะสม เพราะสอดคล้องกับคำแนะนำของ Spector (1992)

อย่างไรก็ตาม การวิเคราะห์ปัจจัยในส่วนของการประเมินความตรงเชิงโครงสร้างที่พิจารณาในส่วนนี้ จะใช้การวิเคราะห์ปัจจัยเชิงยืนยัน แต่พบว่าในขั้นตอนของการวิเคราะห์รายการคำถาม มีการใช้การวิเคราะห์ปัจจัยเชิงสำรวจเพื่อพิจารณาโครงสร้างของตัวแปรด้วย ดังนำเสนอแล้วในหัวข้อ 1.2.6

### 2.2.8 การประเมินความตรงตามเกณฑ์สัมพัทธ์

ความตรงตามเกณฑ์สัมพัทธ์หมายถึง เครื่องมือวัดได้ตรงกับเกณฑ์หรือเครื่องมือวัดมาตรฐาน (gold standard) เมื่อพิจารณาจากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่ามีการรายงานประเด็นการประเมินความตรงตามเกณฑ์สัมพัทธ์ 2/13 เรื่อง ใน 2 เรื่องนี้ รายงานการประเมินความตรงร่วมสมัย (concurrent validity) โดยมีเครื่องมือมาตรฐานในการเปรียบเทียบทั้ง 2 เรื่อง สถิติที่ใช้คือสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เรื่องแรกมีวิธีการประเมินความตรงร่วมสมัยโดยใช้วิธีการคำนวณค่าความสัมพันธ์ระหว่างเครื่องมือมาตรฐานซึ่งใช้เครื่องมือ WHOQOL-BREF เทียบกับสเกลที่สร้างขึ้นคือ Pictorial Thai Quality of Life (PTQL) โดยใช้สถิติ

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สำหรับเรื่องที่สอง ใช้วิธีการคำนวณค่าความสัมพันธ์ระหว่างเครื่องมือมาตรฐานซึ่งใช้ Health Related Self Report Scale (HRSR) เทียบกับสเกลที่สร้างขึ้น โดยใช้สถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเช่นกัน ซึ่งพบการรายงานเฉพาะในบทความวิจัยเท่านั้น สำหรับงานที่เหลือไม่ได้ประเมินความตรงตามเกณฑ์สัมพัทธ์ ทั้งนี้ น่าจะเป็นเพราะว่าเครื่องมือประเมินคุณภาพชีวิตโดยส่วนใหญ่ ไม่มีเครื่องมือมาตรฐานสำหรับการเปรียบเทียบ ดังนั้นจึงเป็นข้อจำกัดในการทดสอบความตรงตามเกณฑ์สัมพัทธ์ของเครื่องมือประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพที่สร้างใหม่ดังที่ Coons (2002) กล่าวไว้

### 2.2.9 การประเมินความเที่ยงของสเกล

เครื่องมือวัดคุณภาพชีวิตจะประกอบคำถามหลาย ๆ ข้อ ดังนั้น ความเที่ยงของเครื่องมือในกรณีนี้ หมายถึง ความเที่ยงประเภทความสอดคล้องภายใน (internal consistency) จากบทความวิจัยและวิทยานิพนธ์ ที่บททวนทั้งหมด 13 เรื่อง พบว่า ทุกงานมีการรายงานค่าความเที่ยงด้วยสถิติสัมประสิทธิ์แอลฟาของครอนบาค ซึ่งสอดคล้องกับการศึกษาของ Hinkin (1995) Hensley (1999) Coste (1995) และ Coste (1997) การรายงานค่าสัมประสิทธิ์แอลฟาของครอนบาคในงานที่บททวนครั้งนี้ มีการรายงานค่าในภาพรวม และรายด้านจำนวน 11/13 เรื่อง และ 9/13 เรื่อง ตามลำดับ ซึ่งเป็นไปตามคำแนะนำของ Spector (1992) ที่กล่าวไว้ว่าการหาความเที่ยงของเครื่องมือโดยคำนวณค่าสัมประสิทธิ์แอลฟา ควรคำนวณรวมทุกด้านและรายด้าน ค่าสัมประสิทธิ์แอลฟาที่รายงานในงานที่บททวนครั้งนี้มีค่าระหว่าง 0.59 - 0.97 โดยมีสองเรื่องที่มีค่าสัมประสิทธิ์แอลฟาต่ำกว่า 0.7 ดังนั้น จึงกล่าวได้ว่า งานที่บททวนทั้งหมดได้ดำเนินการขั้นตอนนี้ซึ่งเป็นขั้นตอนที่จำเป็นต้องทำ และงานส่วนใหญ่ (11/13) ที่บททวนมีค่าความเที่ยงอยู่ในระดับที่ยอมรับ นั่นคือ ค่าสัมประสิทธิ์แอลฟาไม่ควรต่ำกว่า 0.7 แต่ไม่เกิน 0.9 (Nunnally, 1978 อ้างถึงใน Streiner and Norman, 2003)

### 2.2.10 การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง (sensitivity to change)

ความไวต่อการเปลี่ยนแปลงของเครื่องมือ (sensitive to change) คือความสามารถของเครื่องมือในการตรวจจับการเปลี่ยนแปลงของภาวะทางสุขภาพที่เกิดขึ้นในตัวผู้ถูกวัด (Maurice et al., 1998; Fayers and Machin, 2000) การประเมินคุณสมบัติประการนี้ของเครื่องมือ ยังระบุไม่ได้ว่าวิธีการคำนวณหรือสถิติใดดีที่สุด ดังนั้นจึงควรใช้หลายวิธีเพื่อยืนยันผล อีกทั้งควรพิจารณาความสัมพันธ์ของค่าความไวที่คำนวณได้กับมาตรฐานการรักษาอื่น ๆ เช่น การเปลี่ยนแปลงทางคลินิก แล้วจึงพิจารณาความสามารถของเครื่องมือในการวัดความแตกต่างระหว่างกลุ่ม (Fayers and Machin, 2000; Streiner and Norman, 2003) อย่างไรก็ตาม ในหลายสถานการณ์ที่เป็นจริงของคนไข้หรือด้านสุขภาพ ไม่เอื้อต่อการประเมินนี้ ดังนั้น จึงมักพบว่า ในการสร้างเครื่องมือวัดคุณภาพชีวิต

ที่เกี่ยวข้องกับสุขภาพ ไม่มีการประเมินคุณสมบัตินี้ ซึ่งจากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง ไม่พบการรายงานในประเด็นดังกล่าวเช่นกัน

#### 2.2.11 การรายงานการนำไปใช้

จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบ 6/13 เรื่อง รายงานการนำไปใช้จำนวน รายละเอียดที่ระบุคือวิธีคิดคะแนนและการแปลผล โดยพบเฉพาะในวิทยานิพนธ์เท่านั้น ซึ่ง Spector (1992) และ พรณทิพา ศักดิ์ทอง (2550) กล่าวว่าเครื่องมือที่สร้างใหม่ควรมีคู่มือการใช้ มีวิธีการคิดคะแนนและการแปลผล และมีรายละเอียดของหน่วยงานหรือบุคคลที่สามารถติดต่อเพื่อขอ นำเครื่องมือไปใช้ ดังนั้น จึงสรุปว่า ในงานที่ทบทวนยังดำเนินงานส่วนนี้ไม่เป็นไปตามระเบียบวิธีมาตรฐานของการสร้างเครื่องมือ

