

Rater Agreement Index (RAI)

Surachai Meechan

Ed.D. (Testing and Measurement), Lecturer,
Department of Educational Evaluation and Research, Faculty of Education,
Prince of Songkla University
E-mail : msuracha@eduadm.edu.psu.ac.th

Abstract

We can use Rater Agreement Index (RAI) as an index of agreement level of scores gained from two or more raters. This index has a value from 0 to 1. If its value is nearly to 1, it means the raters have a strong agreement in scoring. On the other hand, if its value is nearly to 0, it means that that agreement is not strong. Thus, if the raters are trained and have a good knowledge of scoring, this index can indicate the standard of those scoring rubrics.

Keywords : performance assessment, rater agreement index, scoring rubrics

ดัชนีความสอดคล้องระหว่างผู้ประเมิน

สุรัชย์ มีชาญ

กศ.ด. (การทดสอบและวัดผลการศึกษา), อาจารย์

ภาควิชาประเมินผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์

E- mail : msuracha@eduadm.edu.psu.ac.th

บทคัดย่อ

ดัชนีความสอดคล้องระหว่างผู้ประเมิน (Rater Agreement Index : RAI) เป็นตัวบ่งชี้ถึงระดับความสอดคล้องกันของคะแนนที่ได้จากผู้ประเมิน 2 คน หรือมากกว่า โดยดัชนีนี้จะมีค่าตั้งแต่ 0-1 เมื่อใดที่มีค่าเข้าใกล้ 1 แสดงว่าผู้ประเมินสามารถให้คะแนนได้อย่างสอดคล้องกันสูงมาก แต่ถ้ามีค่าเข้าใกล้ 0 ก็แสดงว่ามีความสอดคล้องกันไม่มากนัก ดังนั้นหากผู้ประเมินคือผู้ที่ได้รับการฝึกฝน และมีความรู้ความเข้าใจเกี่ยวกับเกณฑ์การให้คะแนนอย่างดีแล้ว ดัชนีนี้จะช่วยบ่งชี้ถึงมาตรฐานของเกณฑ์การให้คะแนนดังกล่าวได้

คำสำคัญ : การวัดภาคปฏิบัติ, เกณฑ์การให้คะแนน, ดัชนีความสอดคล้องระหว่างผู้ประเมิน

บทนำ

ทุกครั้งที่จัดให้มีการวัดผลการเรียนรู้ในห้องเรียน ควรจะมีเป้าหมายหลักเพื่อค้นหาจุดเด่น และข้อบกพร่องในการเรียนรู้ของนักเรียนแต่ละคน เพื่อที่ครูจะได้หาแนวทางส่งเสริมและปรับปรุงแก้ไขได้อย่างถูกต้องและสอดคล้องเหมาะสมกับสภาพของแต่ละบุคคลมากที่สุด ดังนั้น สิ่งสำคัญอย่างมากที่ครูควรดำเนินการก็คือการสร้างเครื่องมือวัดผลการเรียนรู้ที่มีคุณภาพและการนำเครื่องมือวัดนั้นไปใช้อย่างถูกต้อง โดยเฉพาะยุคปัจจุบันที่การวัดผลการเรียนรู้ได้เน้นหนักไปที่การวัดภาคปฏิบัติ (performance assessment) ซึ่งมุ่งเน้นไปที่การกระทำหรือการปฏิบัติได้ของนักเรียนแต่ละคนมากยิ่งขึ้น ด้วยเหตุนี้เองประสิทธิภาพของการวัดผลการเรียนรู้จึงขึ้นอยู่กับคุณภาพของแบบสังเกตหรือเครื่องมือวัดอื่น ๆ ที่ต้องใช้ดุลยพินิจของผู้ประเมิน (rater) มากยิ่งขึ้นด้วย เพราะ

เมื่อจะวัดความสามารถในการปฏิบัติก็จำเป็นต้องเกี่ยวข้องกับทักษะในกระบวนการ (process) และคุณภาพของผลงาน (product) ที่นักเรียนได้จัดทำขึ้น ซึ่งจำเป็นต้องอาศัยดุลยพินิจของผู้ประเมินทั้งสิ้น

แต่ปัญหาที่ครูและสถานศึกษาต่าง ๆ ประสบอยู่เสมอก็คือการขาดเครื่องมือวัดภาคปฏิบัติที่มีคุณภาพ โดยเฉพาะเครื่องมือที่มีเกณฑ์การให้คะแนน (scoring rubrics) ที่เหมาะสมและเป็นที่ยอมรับจากทุกฝ่ายที่เกี่ยวข้อง และถึงแม้จะมี ก็พบว่าส่วนใหญ่ไม่ได้ตรวจสอบคุณภาพอย่างเต็มกระบวนการมากนัก โดยเฉพาะการตรวจสอบว่าเครื่องมือเหล่านั้นจะสามารถนำไปใช้ได้โดยตรงและยุติธรรมกับนักเรียนทุกคนหรือไม่ ทั้งนี้ส่วนหนึ่งคงเป็นเพราะไม่ทราบวิธีการ ขณะที่บางส่วนอาจไม่เห็นความจำเป็น แต่หากเรามุ่งหวังให้การวัดและ

ประเมินผลการเรียนรู้ในทุกครั้งมีมาตรฐานเป็นที่ยอมรับจากทุกฝ่าย ก็มีความจำเป็นอย่างยิ่งที่เครื่องมือวัดต่าง ๆ ควรจะได้รับการตรวจสอบคุณภาพอย่างดีที่สุดเท่าที่จะทำได้ก่อนนำไปใช้จริง ทั้งนี้ก็เพื่อให้ผลการวัดที่ถูกต้องและน่าเชื่อถือมากที่สุด

คุณภาพประการหนึ่งที่สำคัญยิ่งของเครื่องมือวัดภาคปฏิบัติ คือ ความชัดเจนและสื่อความได้ตรงกัน ซึ่งเรียกกันว่า ความเป็นปรนัย (objectivity) โดยเฉพาะในส่วนของการกำหนดการให้คะแนน ซึ่งจำเป็นต้องมีมาตรฐานอย่างมาก นั่นคือ ไม่ว่าจะนำไปใช้เมื่อใดหรือใครคนใดจะเป็นผู้ใช้ก็ตาม คะแนนเหล่านั้นควรจะคงที่หรือใกล้เคียงกันมากที่สุดเท่าที่จะเป็นไปได้ ซึ่งคุณภาพดังกล่าวนี้ นับว่าเป็นสิ่งที่อ่อนด้อยอย่างมากในเครื่องมือวัดทุกประเภทที่ต้องใช้ดุลยพินิจในการตัดสินใจของครูหรือผู้ประเมิน แม้แต่ข้อสอบชนิดบรรยายหรือความเรียง (essay item) ก็ไม่พ้นที่จะประสบปัญหานี้เช่นเดียวกัน โดยเฉพาะในส่วนของการวัดภาคปฏิบัตินั้นไม่ว่าจะมุ่งเน้นไปที่กระบวนการหรือผลงานก็ล้วนมีจุดอ่อนนี้ทั้งสิ้น ซึ่งไม่ว่าจะเกิดขึ้นจากอคติหรือความไม่รู้จริงของผู้ประเมินหรือเกิดจากความไม่ชัดเจนของการให้คะแนนที่กำหนดไว้ก็ตาม จุดอ่อนนี้ควรจะได้รับแก้ไขให้หมดไปหรือเหลือน้อยที่สุดเท่าที่ครูจะพึงกระทำได้

การกำหนดเกณฑ์การให้คะแนน

โดยทั่วไป เครื่องมือวัดภาคปฏิบัติจะประกอบด้วย 2 ส่วนใหญ่ ๆ คือ ส่วนที่เป็นแบบบันทึกคะแนน ซึ่งครูจะต้องใช้ในการจดบันทึกคะแนนและข้อสังเกตต่าง ๆ กับส่วนที่เป็นเกณฑ์การให้คะแนน เพื่ออธิบายให้ครูทราบว่าจะต้องพิจารณาประเด็นใดบ้างและให้คะแนนในแต่ละประเด็นอย่างไร ซึ่งสองส่วนนี้อาจพิมพ์ไว้คนละด้านหรืออาจจัดพิมพ์แยกไว้ต่างหากก็แล้วแต่ความเหมาะสม ในที่นี้จะขอกกล่าวถึงเฉพาะลักษณะและตัวอย่างของเกณฑ์การให้คะแนนที่มีผู้นำเสนอไว้แล้วอย่างคร่าว ๆ เพื่อเป็นการเกริ่นนำเข้าสู่ขั้นตอนการตรวจสอบคุณภาพของเกณฑ์ดังกล่าวอีกต่อหนึ่งเท่านั้น โดยจะไม่กล่าวถึงลักษณะและตัวอย่างของแบบบันทึกคะแนนแต่ประการใด ทั้งนี้เพราะสามารถศึกษาค้นคว้าในประเด็นเหล่านี้ได้จากเอกสารตำราทางด้าน การวัด

และประเมินผลการศึกษาที่พิมพ์เผยแพร่อยู่ทั่วไป และตามแนวทางที่ปฏิบัติกันโดยทั่วไปนั้น ผู้สร้างเครื่องมือวัดก็มีอิสระที่จะกำหนดรูปลักษณ์ของแบบบันทึกคะแนนได้อย่างหลากหลาย ขึ้นอยู่กับความพึงพอใจและความสะดวกในการนำไปใช้เป็นสำคัญ

รุ่งศิริรินทร์ จันทรหอม (2545, 26) ได้นำเสนอตัวอย่างเกณฑ์การให้คะแนน สำหรับใช้ในการวัดภาคปฏิบัติ ทั้งในส่วนของการกำหนดเกณฑ์การให้คะแนนแบบภาพรวม (holistic scoring rubrics) อันหมายถึงการให้คะแนนโดยพิจารณาคุณภาพของผลงานหรือการกระทำในภาพรวมทั้งหมด และการกำหนดเกณฑ์การให้คะแนนแบบแยกองค์ประกอบ (analytic scoring rubrics) ซึ่งเป็นการให้คะแนนโดยการแบ่งย่อยกระบวนการหรือผลงานนั้นออกเป็น ส่วน ๆ แล้วจึงพิจารณาให้คะแนนในแต่ละส่วนนั้น ดังปรากฏในตัวอย่างที่ 1 และ 2 ดังนี้

ตัวอย่างที่ 1

เกณฑ์การให้คะแนนแบบภาพรวม สำหรับงานประดิษฐ์ใบตองกลีบบัวหลวง

4 (ดีมาก) หมายถึง เตรียมวัสดุอุปกรณ์ครบถ้วน เหมาะสม ฉีกใบตองเท่ากันทุกชิ้น ไม่ซ้ำ พับกลีบปลายแหลมสมบูรณ์ทุกกลีบ การส่งกลีบเสมอกันทุกกลีบ รอยเย็บเรียบแข็งแรงและสามารถนำไปใช้ได้ตามวัตถุประสงค์

3 (ดี) หมายถึง เตรียมวัสดุอุปกรณ์ครบถ้วน เหมาะสม ฉีกใบตองเท่ากันทุกชิ้น ไม่ซ้ำ พับกลีบปลายแหลมสมบูรณ์บางกลีบ การส่งกลีบเสมอกันเกือบทุกกลีบ รอยเย็บเรียบ แข็งแรงพอใช้ สามารถนำไปใช้ได้ตามวัตถุประสงค์

2 (พอใช้) หมายถึง เตรียมวัสดุอุปกรณ์มาบางส่วน ฉีกใบตองไม่ซ้ำแต่ยังไม่เท่ากันทุกชิ้น พับกลีบปลายไม่แหลม การส่งกลีบไม่เสมอกัน รอยเย็บไม่เรียบนำไปใช้งานยังไม่เป็นที่น่าพอใจนัก

1 (ควรปรับปรุง) หมายถึง ยี่วัสดุอุปกรณ์จากเพื่อน ปฏิบัติงานไม่ตลอดเวลา ผลงานไม่สำเร็จ

ตัวอย่างที่ 2

เกณฑ์การให้คะแนนแบบแยกองค์ประกอบ สำหรับงานประดิษฐ์สิ่งของจากวัสดุเหลือใช้

รายการ	ระดับคุณภาพ		
	3 (ดี)	2 (พอใช้)	1 (ควรปรับปรุง)
การเตรียมงาน	สะสมวัสดุเหลือใช้ได้พอเพียงและเหมาะสม	สะสมวัสดุเหลือใช้ได้พอเพียงและเหมาะสมบางส่วน	สะสมวัสดุเหลือใช้ได้ไม่พอเพียง
ลำดับขั้นในการทำงาน	ปฏิบัติงานได้ตามลำดับขั้นตอน	ปฏิบัติงานเป็นไปตามขั้นตอนบางส่วน	ปฏิบัติงานไม่เป็นไปตามขั้นตอน
ความคิดริเริ่มสร้างสรรค์	ความคิด ผลงาน ผู้เรียนคิดขึ้นเอง	ความคิด ผลงาน ผู้เรียนค้นคว้า ลอกเลียนแบบจากหนังสือ	ความคิด ผลงาน ลอกเลียนแบบจากเพื่อน
ใช้งานได้ตามวัตถุประสงค์	ผลงานนำไปใช้ได้ตามวัตถุประสงค์	ผลงานนำไปใช้ยังไม่เป็นที่น่าพอใจนัก	ผลงานไม่เสร็จสมบูรณ์
ความรับผิดชอบ	ทำงานที่ได้รับมอบหมายครบถ้วน และส่งงานตามกำหนดเวลา	ทำงานที่ได้รับมอบหมายครบถ้วน แต่ส่งงานช้ากว่ากำหนดเวลา	ทำงานที่ได้รับมอบหมายบางครั้ง และส่งงานที่ไม่เสร็จสมบูรณ์
ความประณีต เรียบร้อย	ผลงานมีความประณีต เรียบร้อยดี	ผลงานเรียบร้อยพอใช้	ผลงานขาดความประณีต เรียบร้อย
ความมีน้ำใจ	ช่วยเหลือผู้อื่นสม่ำเสมอ	ช่วยเหลือผู้อื่นบ้างในบางครั้ง	หลีกเลี่ยงการช่วยเหลือผู้อื่น

ตัวอย่างที่ 3

แบบวัดฉบับหนึ่งถูกสร้างขึ้นเพื่อวัดทักษะของนักเรียนในการพูดสื่อสารที่หน้าชั้นเรียน โดยทักษะดังกล่าวประกอบด้วยพฤติกรรมบ่งชี้หรือองค์ประกอบที่สำคัญรวม 4 ประการ และกำหนดเกณฑ์การให้คะแนนไว้ดังต่อไปนี้ (ดัดแปลงจาก สุวิมล ว่องวานิช, 2535, 516)

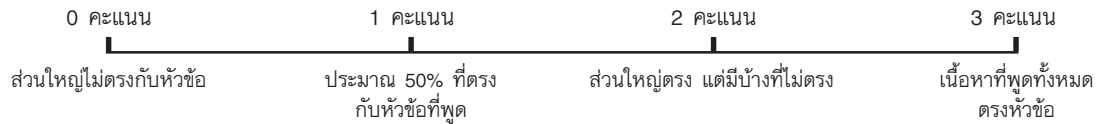
จากตัวอย่างเกณฑ์การให้คะแนนและปัญหาด้านคุณภาพตามที่ได้กล่าวมาแล้วนั้น แนวทางหนึ่งที่ครูควรดำเนินการก็คือการตรวจสอบว่าเครื่องมือวัด

ฉบับนั้นจะสามารถนำไปใช้ได้อย่างมีมาตรฐานมากน้อยเพียงใด ทั้งนี้โดยการมอบหมายให้ผู้ประเมินอย่างน้อย 2 คนได้ทดลองประเมินการทำงานหรือผลงานของนักเรียนหนึ่งคนหรือมากกว่า แล้วจึงใช้กระบวนการทางสถิติมาช่วยตรวจสอบความสอดคล้องของคะแนนที่ได้จากผู้ประเมินเหล่านั้นว่ามีมากน้อยเพียงใด ถ้าพบว่ามี ความสอดคล้องกันอย่างมาก ย่อมบ่งชี้ว่าเครื่องมือวัดฉบับนั้นมีมาตรฐานเพียงพอที่จะนำไปใช้ต่อไป และหนึ่งในวิธีการทางสถิติที่น่าจะนำมาปรับใช้ได้อย่างดียิ่งก็คือดัชนีความสอดคล้องระหว่าง

เกณฑ์การให้คะแนน

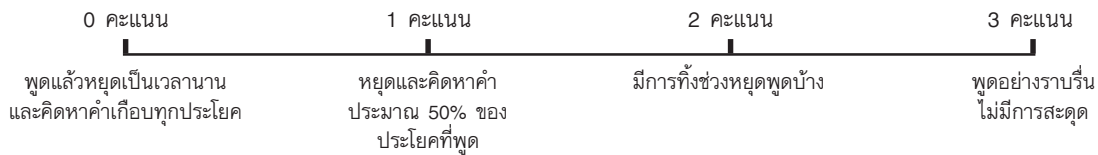
1. ความตรงของเนื้อหา

เนื้อหาที่พูดตรงกับหัวข้อหรือไม่ ?



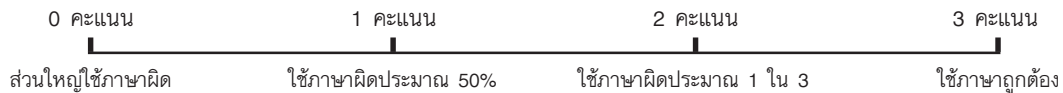
2. ความต่อเนื่องของการพูด

การพูดดำเนินไปอย่างราบรื่นหรือไม่ ?



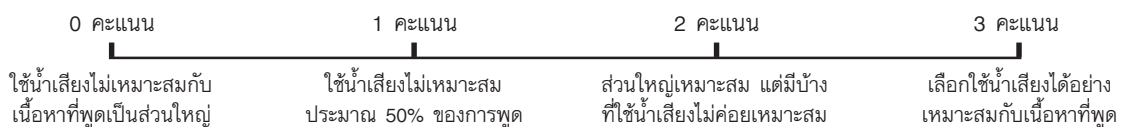
3. ความถูกต้องในการใช้ภาษา

ใช้ภาษาได้ถูกต้องหรือไม่ ?



4. ความเหมาะสมของน้ำเสียง

ใช้น้ำเสียงเหมาะสมหรือไม่ ?



ผู้ประเมินซึ่งนำเสนอไว้โดย Judith A. Burry-Stock และคณะ (Burry-Stock & other, 1996)

ดัชนีความสอดคล้องระหว่างผู้ประเมิน

เราสามารถหาดัชนีความสอดคล้องระหว่างผู้ประเมินเพื่อตรวจสอบคุณภาพของเครื่องมือวัดที่กำหนดคะแนนไว้เป็นช่วง ๆ ได้อย่างหลากหลาย ทั้งนี้ให้คะแนนแบบ 0-1 หรือแบบหลายค่าต่าง ๆ เช่น ให้

0-1-2-3, 1-2-3-4 หรืออื่น ๆ ซึ่งในการคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินที่จะกล่าวถึงต่อไปนี้ ผู้เขียนจะได้นำเสนอแยกเป็นประเด็นย่อย ๆ ตามจำนวนของพฤติกรรมบ่งชี้หรือองค์ประกอบที่มุ่งวัด จำนวนของนักเรียนหรือผลงาน(ในกรณีที่ต้องการประเมินผลงาน) และจำนวนของผู้ประเมินทั้งหมด โดยมีการเพิ่มเติมตัวแปรและสัญลักษณ์ในสูตรคำนวณเดิมบ้างเล็กน้อย ทั้งนี้ด้วยความคาดหวังว่าจะช่วยให้สื่อความเข้าใจได้ดี

ยิ่งขึ้น และทุกครั้งที่มีการกล่าวถึงดัชนีความสอดคล้องระหว่างผู้ประเมินที่คำนวณมาได้จะหมายถึงเฉพาะกรณี que ผู้ประเมินทุกคนได้รับคำชี้แจงและฝึกฝนการใช้เครื่องมือวัดอย่างดีที่สุดแล้ว ทั้งนี้เพื่อให้ดัชนีดังกล่าวสามารถบ่งชี้ไปถึงคุณภาพของเครื่องมือวัดอย่างแท้จริง เพราะถ้าไม่ใช่ ดัชนีนี้อาจจะบ่งชี้กลับไปยังมาตรฐานของผู้ประเมินแทน (แต่ก็คงไม่ใช่เรื่องยากนัก หากจะลองเชื่อมโยงความคิดจากประเด็นแรกไปสู่ประเด็นที่สองหรือประเด็นอื่น ๆ ต่อไปอีก) ดังรายละเอียดต่อไปนี้

1. กรณีที่มีพฤติกรรมบ่งชี้เพียงหนึ่งตัว นักเรียนคนเดียว และมีผู้ประเมิน 2 คน

ในกรณีที่ทดลองใช้เครื่องมือวัดนั้นกับนักเรียนเพียงคนเดียวโดยมีพฤติกรรมบ่งชี้เพียงหนึ่งตัวและมีผู้ประเมิน 2 คน คะแนนที่ได้จากผู้ประเมินแต่ละคนจึงมีเพียง 1 ค่าเช่นเดียวกัน ดังนั้นในการคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน 2 คนใด ๆ จึงสามารถใช้แนวคิดพื้นฐานอย่างง่ายได้ว่า หากผู้ประเมินทั้งสองให้คะแนนเท่ากันแล้ว "ผลต่างของคะแนนที่ได้จากผู้ประเมินทั้งสอง" จะมีค่าเท่ากับ 0 ซึ่งจะส่งผลให้สัดส่วน (proportion) ของ "ผลต่างของคะแนนที่ได้จากผู้ประเมินทั้งสอง" กับ "ผลต่างสูงสุดของคะแนนที่เป็นไปได้" มีค่าเท่ากับ 0 ด้วย (มีความสอดคล้องกันอย่างสมบูรณ์) แต่เมื่อใดที่ผู้ประเมินทั้งสองให้คะแนนต่างกันมากขึ้น สัดส่วนดังกล่าวก็จะมีค่าสูงขึ้นตามไปด้วย โดยจะมีค่าเท่ากับ 1 เมื่อคะแนนทั้งสองต่างกันสุดขีด ดังนั้นหากกำหนดให้ I แทน จำนวนของคะแนนที่เป็นไปได้ทั้งหมด เช่น เมื่อกำหนดคะแนนไว้เป็น "0-1-2-3 คะแนน" กรณีนี้จะมีจำนวนของคะแนนที่เป็นไปได้ เท่ากับ 4 ค่า (I = 4) ดังนั้นผลต่างสูงสุดของคะแนนที่เป็นไปได้จากผู้ประเมิน 2 คนใด ๆ จึงมีค่าเท่ากับ I - 1 หรือ 4 - 1 หรือ 3 คะแนน ดังนั้นหากผู้ประเมินทั้งสองให้คะแนนต่างกันสุดขีดคือ "0 กับ 3" ย่อมสะท้อนให้เห็นว่าไม่มีความสอดคล้องระหว่างผู้ประเมินทั้งสองคนนั้น ในทำนองกลับกันหากผู้ประเมินทั้งสองให้คะแนนเท่ากันก็แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินอย่างสมบูรณ์จากหลักการพื้นฐานนี้จึงอาจกล่าวอย่างรวบรัดถึงสูตรที่ใช้ในการคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน (2 คน) ได้ดังนี้

$$RAI = 1 - \frac{|R_1 - R_2|}{I - 1} \dots\dots\dots 1$$

โดยที่

- RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน
- R₁ แทน คะแนนที่ได้จากผู้ประเมินคนที่ 1
- R₂ แทน คะแนนที่ได้จากผู้ประเมินคนที่ 2
- I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้

(ตามเกณฑ์การให้คะแนน)

ทั้งนี้ดัชนีความสอดคล้องระหว่างผู้ประเมินจะมีค่าสูงสุดเท่ากับ 1 เมื่อผู้ประเมินทั้งสองคนให้คะแนนเท่ากัน และมีค่าต่ำสุดเท่ากับ 0 เมื่อผู้ประเมิน 2 คนนั้นให้คะแนนต่างกันสุดโต่งหรือสวนทางกันอย่างสิ้นเชิง กล่าวคือหากคนหนึ่งให้คะแนนสูงสุดเท่าที่จะเป็นไปได้ อีกคนหนึ่งจะให้คะแนนต่ำสุดเท่าที่จะเป็นไปได้ ซึ่งจะทำให้ผลต่างของคะแนนที่ได้จากผู้ประเมินทั้งสอง (ค่าสัมบูรณ์ของ R₁-R₂ : |R₁-R₂|) มีค่าเท่ากับ I - 1 อันจะส่งผลให้ RAI มีค่าเท่ากับ 0 ดังกล่าวแล้ว

อย่างไรก็ตามในการนำไปประยุกต์ใช้ก็ไม่ได้มีการกำหนดเกณฑ์ไว้อย่างเด่นชัดว่าดัชนีความสอดคล้องระหว่างผู้ประเมินที่คำนวณได้ควรมีค่าเท่าใดจึงจะถือว่ามีความสอดคล้องระหว่างผู้ประเมินมากพอที่จะยอมรับ เพราะภายใต้เงื่อนไขของการปฏิบัติจริง เราจำเป็นต้องพิจารณาตัดสินใจด้วยว่าจากจุดประสงค์ของการใช้เครื่องมือวัดในครั้งนั้น ๆ จำเป็นต้องใช้เครื่องมือที่มีคุณภาพมากน้อยเพียงใด ดังนั้นการแปลความหมายของค่าดัชนีความสอดคล้องระหว่างผู้ประเมินที่ปรากฏในบทความนี้จึงเป็นการใช้ค่าตามความหมายอย่างกว้าง ๆ เพียงเพื่อให้สื่อความกันโดยทั่วไปเท่านั้น

ตัวอย่าง

จากการประเมินงานประดิษฐ์ใบตองกลีบบัวหลวงของนักเรียนคนหนึ่งโดยผู้ประเมิน 2 คน ด้วยเครื่องมือวัดที่มีเกณฑ์การให้คะแนน รวม 4 ค่า คือ 1-2-3-4 คะแนน (จากตัวอย่างที่ 1)

สมมติว่า ผู้ประเมินคนแรกให้ 4 คะแนน และผู้ประเมินคนที่สองให้เพียง 2 คะแนน ผลต่างของคะแนนที่ได้จากผู้ประเมินทั้งสอง หรือ |R₁-R₂| จะมี

ค่าเท่ากับ 2 คะแนน ขณะที่ผลต่างสูงสุดที่เป็นไปได้เท่ากับ 1 - 1 หรือ 4 - 1 หรือ 3 คะแนน ดังนั้น RAI จะมีค่าเท่ากับ 0.33 ซึ่งค่าดังกล่าวนี้แสดงให้เห็นว่าการตัดสินครั้งนี้ไม่มีมาตรฐานพอที่จะยอมรับได้ เพราะแม้จะมีเกณฑ์การให้คะแนนที่กำหนดไว้ร่วมกันแล้ว แต่ผู้ประเมินทั้งสองก็ยังไม่สามารถให้คะแนนอย่างสอดคล้องกันได้ ทั้งนี้จึงมีความจำเป็นต้องพิจารณาแก้ไขเกณฑ์การให้คะแนนให้มีคุณภาพต่อไป

ในกรณีที่กำหนดคะแนนไว้เพียง 2 ค่า (I = 2) เช่น เมื่อให้ 0-1 หรือ 1-2 คะแนน จะทำให้ 1 - 1 มีค่าเท่ากับ 1 จึงสามารถเขียนสูตรใหม่ให้ง่ายขึ้นกว่าเดิมได้ดังนี้

$$RAI = 1 - |R_1 - R_2| \dots\dots\dots 2$$

แม้ว่าโอกาสที่จะได้ใช้สูตรที่ 1 และ 2 คงมีไม่บ่อยนัก เพราะเครื่องมือวัดที่เราสร้างขึ้นมักจะถูกนำไปทดลองใช้กับนักเรียนจำนวนมาก และโดยทั่วไปก็มักจะมีพฤติกรรมบ่งชี้หลาย ๆ ตัว แต่สาเหตุที่นำสูตรนี้มากล่าวถึงไว้ด้วยก็เพื่อให้ทราบแนวคิดพื้นฐานของการพัฒนาสูตรต่าง ๆ ที่ซับซ้อนขึ้นดังที่ได้กล่าวถึงต่อไป

2. กรณีที่มีพฤติกรรมบ่งชี้เพียงหนึ่งตัว นักเรียนคนเดียว และมีผู้ประเมินหลายคน

ในกรณีที่น่าเครื่องมือวัดที่มีพฤติกรรมบ่งชี้เพียงหนึ่งตัวไปทดลองใช้กับนักเรียนเพียงคนเดียว แต่กำหนดให้มีผู้ประเมินหลายคน เราสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินได้ด้วยสูตรดังต่อไปนี้

$$RAI = 1 - \frac{\sum_{m=1}^M |R_m - \bar{R}|}{(M-1)(I-1)} \dots\dots\dots 3$$

โดยที่

RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน
R_m แทน คะแนนที่ได้จากผู้ประเมินคนที่ m (m = 1, 2, 3,.....,M)

\bar{R} แทน คะแนนเฉลี่ย หรือค่าเฉลี่ยเลขคณิต (arithmetic mean) ของคะแนนที่ได้จากผู้ประเมินทุกคน ซึ่งคำนวณได้จากสูตร

$$\bar{R} = \frac{\sum_{m=1}^M R_m}{M}$$

M แทน จำนวนของผู้ประเมินทั้งหมด

I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้ (ตามเกณฑ์การให้คะแนน)

ทั้งนี้หากพิจารณาโดยอาศัยหลักสถิติจะเห็นได้

ว่าผลลัพธ์ของคะแนนแต่ละตัวลดด้วยค่าเฉลี่ย (R_m - \bar{R}) ก็คือ คะแนนเบี่ยงเบน (deviate score) นั่นเอง ซึ่งหากผลรวมของค่าสัมบูรณ์ของคะแนนเบี่ยงเบนทั้งหมดมีค่ามากย่อมแสดงว่าผู้ประเมินแต่ละคนให้คะแนนที่แตกต่างกันมาก ในทางตรงกันข้ามถ้าผลรวมดังกล่าวมีค่าลดลงก็แสดงว่าผู้ประเมินให้คะแนนได้อย่างสอดคล้องกันมากขึ้นด้วย โดยเมื่อใดที่ผู้ประเมินทุกคนสามารถให้คะแนนได้เท่ากัน คะแนนเบี่ยงเบนทุกตัวก็จะมีค่าเท่ากับ 0 ซึ่งจะทำให้ RAI มีค่าเท่ากับ 1 ดังนั้นหากเราย้อนกลับไปพิจารณาสูตรที่ 1 อีกครั้งจะพบว่าเป็นสูตรเดียวกับสูตรที่ 3 นั่นเอง เพียงแต่มีการปรับเปลี่ยนรูปสมการใหม่ เพราะในกรณีที่มีผู้ประเมิน 2 คน (M = 2) ค่าของ M - 1 จะเท่ากับ 1 จึงไม่ปรากฏ M - 1 อยู่ในสูตรดังกล่าว และ

ในทางสถิตินั้น $\sum_{m=1}^2 |R_m - \bar{R}|$ ย่อมมีค่าเท่ากับ $|R_1 - R_2|$

จึงสามารถใช้แทนกันได้เสมอ ซึ่งในกรณีนี้ท่านสามารถพิจารณาถึงความเท่ากันของทั้งคู่ได้ด้วยเส้นกราฟต่อไปนี้



ตัวอย่าง

สมมติว่าคะแนนงานประดิษฐ์ภาคใบตองกล้วยบัวหลวงของนักเรียนคนหนึ่งที่ได้จากผู้ประเมิน 3 คน มีค่าเท่ากับ 3, 3 และ 2 คะแนน ตามลำดับ เราสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินได้ดังนี้

R_1	R_2	R_3	\bar{R}	$ R_1 - \bar{R} $	$ R_2 - \bar{R} $	$ R_3 - \bar{R} $
3	3	2	2.67	0.33	0.33	0.67
				$\sum_{m=1}^3 R_m - \bar{R} = 1.33$		

แทนค่าต่าง ๆ ลงในสูตรคำนวณ จะได้

$$RAI = 1 - \frac{\sum_{m=1}^M |R_m - \bar{R}|}{(M-1)(I-1)}$$

$$= 1 - \frac{1.33}{(3-1)(4-1)}$$

$$= 0.78$$

$$RAI = 1 - \frac{\sum_{k=1}^K |R_{1k} - R_{2k}|}{K(I-1)} \dots\dots\dots 4$$

โดยที่

- RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน
- R_{1k} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 1 ในพฤติกรรมที่ k (k = 1, 2, 3,...,K)
- R_{2k} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 2 ในพฤติกรรมที่ k (k = 1, 2, 3,...,K)
- K แทน จำนวนของพฤติกรรมบ่งชี้ทั้งหมด
- I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้ (ตามเกณฑ์การให้คะแนน)

ดัชนีความสอดคล้องระหว่างผู้ประเมินจึงมีค่าเท่ากับ 0.78 ซึ่งแสดงให้เห็นว่าเครื่องมือวัดฉบับนี้มีคุณภาพอยู่ในขั้นใช้ได้ แต่ถ้าสามารถปรับปรุงเกณฑ์การให้คะแนนให้ชัดเจนและสื่อความได้ตรงกันมากขึ้นกว่านี้ก็จะได้มาก

3. กรณีที่มีพฤติกรรมบ่งชี้หลายตัว นักเรียนคนเดียว และมีผู้ประเมิน 2 คน

ในกรณีที่น่าเครื่องมือวัดที่มีพฤติกรรมบ่งชี้หลายตัวไปทดลองใช้กับนักเรียนเพียงคนเดียว โดยกำหนดให้ผู้ประเมิน 2 คน คะแนนที่ได้จากผู้ประเมินแต่ละคนก็จะมีหลายค่าตามจำนวนของพฤติกรรมบ่งชี้ทั้งหมด และสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินได้ด้วยสูตรดังต่อไปนี้

กรณีนี้ RAI จะมีค่าสูงสุดเท่ากับ 1 เมื่อผู้ประเมินทั้งสองสามารถให้คะแนนได้ตรงกันในทุกพฤติกรรม และจะมีค่าต่ำสุดเท่ากับ 0 เมื่อผู้ประเมินทั้งสองคนนั้นให้คะแนนต่างกันสุดโต่งหรือสวนทางกันอย่างสิ้นเชิงในทุกพฤติกรรม

ตัวอย่าง

คะแนนทักษะในการพูดหน้าชั้น (ตามที่ได้กล่าวถึงมาแล้วในตัวอย่างที่ 3) ของนักเรียนคนหนึ่งที่ได้จากผู้ประเมิน 2 คน ปรากฏผลดังนี้

รายการประเมิน	คะแนนที่ได้จากผู้ประเมินคนที่ 1	คะแนนที่ได้จากผู้ประเมินคนที่ 2
1. ความตรงของเนื้อหา	2	3
2. ความต่อเนื่องของการพูด	1	1
3. ความถูกต้องในการใช้ภาษา	3	2
4. ความเหมาะสมของน้ำเสียง	2	2

จากคะแนนที่ปรากฏ สามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน ได้ดังนี้

k	R _{1k}	R _{2k}	R _{1k} -R _{2k}
1	2	3	1
2	1	1	0
3	3	2	1
4	2	2	0
			$\sum_{k=1}^4 R_{1k}-R_{2k} = 2$

แทนค่าต่างๆ ลงในสูตรคำนวณ จะได้

$$\begin{aligned} RAI &= 1 - \frac{\sum_{k=1}^K |R_{1k}-R_{2k}|}{K(I-1)} \\ &= 1 - \frac{2}{4(4-1)} \\ &= 0.83 \end{aligned}$$

ดัชนีความสอดคล้องระหว่างผู้ประเมินจึงมีค่าเท่ากับ 0.83 ซึ่งแสดงให้เห็นว่าแบบวัดทักษะการพูดหน้าชั้นฉบับนี้มีคุณภาพอยู่ในขั้นใช้ได้ แต่หากปรับปรุงเกณฑ์การให้คะแนนในพฤติกรรมที่ 1 และ 3 ให้ชัดเจนจนสามารถสื่อความได้ตรงกันมากกว่านี้ก็จะดีมาก

4. กรณีที่มีพฤติกรรมบ่งชี้หลายตัว นักเรียนคนเดียว และมีผู้ประเมินหลายคน

ในกรณีที่นำเครื่องมือวัดที่มีพฤติกรรมบ่งชี้หลายตัวไปทดลองใช้กับนักเรียนเพียงคนเดียว แต่กำหนดให้มีผู้ประเมินหลายคน จะสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินได้ด้วยสูตรดังต่อไปนี้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{m=1}^M |R_{mk} - \bar{R}_k|}{K(M-1)(I-1)} \dots\dots\dots 5$$

โดยที่

RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน
R_{mk} แทน คะแนนที่ได้จากผู้ประเมินคนที่ m ในพฤติกรรมที่ k (m = 1, 2, 3,..., M และ k = 1, 2, 3,...,K)

\bar{R}_k แทน คะแนนเฉลี่ยในพฤติกรรมที่ k ซึ่งคำนวณได้จากสูตร

$$\bar{R}_k = \frac{\sum_{m=1}^M R_{mk}}{M}$$

K แทน จำนวนของพฤติกรรมบ่งชี้ทั้งหมด

M แทน จำนวนของผู้ประเมินทั้งหมด

I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้ (ตามเกณฑ์การให้คะแนน)

ตัวอย่าง

คะแนนทักษะการพูดหน้าชั้นของนักเรียนคนหนึ่งที่ได้จากผู้ประเมิน 3 คน ปรากฏผลดังนี้

รายการประเมิน	คะแนนที่ได้จาก ผู้ประเมินคนที่ 1	คะแนนที่ได้จาก ผู้ประเมินคนที่ 2	คะแนนที่ได้จาก ผู้ประเมินคนที่ 3
1. ความตรงของเนื้อหา	2	3	2
2. ความต่อเนื่องของการพูด	1	1	1
3. ความถูกต้องในการใช้ภาษา	3	2	3
4. ความเหมาะสมของน้ำเสียง	2	2	2

จากคะแนนที่ปรากฏ สามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน ได้ดังนี้

k	R _{1k}	R _{2k}	R _{3k}	\bar{R}_k	$ R_{1k}-\bar{R}_k $	$ R_{2k}-\bar{R}_k $	$ R_{3k}-\bar{R}_k $	$\sum_{m=1}^3 R_{mk}-\bar{R}_k $
1	2	3	2	2.33	0.33	0.67	0.33	1.33
2	1	1	1	1	0	0	0	0
3	3	2	3	2.67	0.33	0.67	0.33	1.33
4	2	2	2	2	0	0	0	0
								$\sum_{k=1}^4 \sum_{m=1}^3 R_{mk}-\bar{R}_k = 2.66$

แทนค่าต่าง ๆ ลงในสูตรคำนวณ จะได้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{m=1}^M |R_{mk} - \bar{R}_k|}{K(M-1)(I-1)}$$

$$= 1 - \frac{2.66}{4(3-1)(4-1)}$$

$$= 0.89$$

ดัชนีความสอดคล้องระหว่างผู้ประเมินจึงมีค่าเท่ากับ 0.89 ซึ่งแสดงให้เห็นว่าแบบวัดทักษะการพูดหน้าชั้นฉบับนี้มีคุณภาพอยู่ในขั้นใช้ได้ แต่หากปรับปรุงเกณฑ์การให้คะแนนในพฤติกรรมที่ 1 และ 3 ให้ชัดเจนจนสามารถสื่อความได้ตรงกันมากกว่านี้ก็จะมีค่า

5. กรณีที่มีพฤติกรรมบ่งชี้หลายตัว นักเรียนหลายคน และมีผู้ประเมิน 2 คน

ในกรณีที่เรานำเครื่องมือวัดที่มีพฤติกรรมบ่งชี้หลายตัวไปทดลองใช้กับนักเรียนกลุ่มหนึ่ง (แทนที่จะทดลองใช้กับนักเรียนเพียงคนเดียว) โดยกำหนดให้มีผู้ประเมิน 2 คน จะสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน ได้ด้วยสูตรดังต่อไปนี้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N |R_{1nk} - R_{2nk}|}{KN(I-1)} \dots\dots\dots 6$$

โดยที่

RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน
 R_{1nk} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 1 ของนักเรียนคนที่ n ในพฤติกรรมที่ k (n = 1, 2, 3,..., N และ k = 1, 2, 3,...,K)

R_{2nk} แทน คะแนนที่ได้จากผู้ประเมินคนที่ 2 ของนักเรียนคนที่ n ในพฤติกรรมที่ k (n = 1, 2, 3,..., N และ k = 1, 2, 3,...,K)

K แทน จำนวนของพฤติกรรมบ่งชี้ทั้งหมด
N แทน จำนวนของนักเรียนทั้งหมด
I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้
(ตามเกณฑ์การให้คะแนน)

ตัวอย่าง
คะแนนทักษะการพูดหน้าชั้นของนักเรียน 5 คน
ที่ได้จากผู้ประเมิน 2 คน ปรากฏผลดังนี้

นักเรียน คนที่	คะแนนที่ได้จากผู้ประเมินคนที่ 1				คะแนนที่ได้จากผู้ประเมินคนที่ 2			
	ความตรง ของเนื้อหา	ความต่อเนื่อง ของการพูด	ความถูกต้อง ในการใช้ภาษา	ความเหมาะสม ของน้ำเสียง	ความตรง ของเนื้อหา	ความต่อเนื่อง ของการพูด	ความถูกต้อง ในการใช้ภาษา	ความเหมาะสม ของน้ำเสียง
1	2	1	3	2	3	1	2	2
2	1	2	3	1	3	2	1	2
3	1	2	2	3	1	2	2	3
4	2	1	3	1	3	2	1	2
5	0	2	3	2	1	2	3	2

จากคะแนนที่ปรากฏ สามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมิน ได้ดังนี้

n \ k	R_{1nk}				R_{2nk}				$ R_{1nk} - R_{2nk} $			
	1	2	3	4	1	2	3	4	1	2	3	4
1	2	1	3	2	3	1	2	2	1	0	1	0
2	1	2	3	1	3	2	1	2	2	0	2	1
3	1	2	2	3	1	2	2	3	0	0	0	0
4	2	1	3	1	3	2	1	2	1	1	2	1
5	0	2	3	2	1	2	3	2	1	0	0	0

$$\sum_{k=1}^4 \sum_{n=1}^5 |R_{1nk} - R_{2nk}| = 13$$

แทนค่าต่าง ๆ ลงในสูตรคำนวณ จะได้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N |R_{1nk} - R_{2nk}|}{KN(I-1)}$$

$$= 1 - \frac{13}{(4)(5)(4-1)}$$

$$= 0.78$$

ดัชนีความสอดคล้องระหว่างผู้ประเมินจึงมีค่าเท่ากับ 0.78 ซึ่งแสดงให้เห็นว่าแบบวัดทักษะการพูดหน้าชั้นฉบับนี้มีคุณภาพอยู่ในขั้นใช้ได้ แต่หากปรับปรุงเกณฑ์การให้คะแนนในบางพฤติกรรมให้ชัดเจนจนสามารถสื่อความได้ตรงกันมากกว่านี้ก็จะดีมาก (จากตัวอย่างนี้พฤติกรรมที่ประเมินได้ตรงกันมากที่สุดคือพฤติกรรมที่ 2)

6. กรณีทั่วไป : เมื่อมีพฤติกรรมบ่งชี้หลายตัว นักเรียนหลายคน และมีผู้ประเมินหลายคน

ในกรณีทั่วไป หากเราต้องการตรวจสอบคุณภาพของเครื่องมือวัดที่มีพฤติกรรมบ่งชี้หลายตัวโดยการนำปทดลองใช้กับนักเรียนกลุ่มหนึ่งและกำหนดให้มีผู้ประเมินหลายคนนั้น เราสามารถคำนวณหาดัชนีความสอดคล้องระหว่างผู้ประเมินได้โดยอาศัยแนวคิดพื้นฐานตามที่ได้กล่าวมาแล้วทั้งหมด ซึ่งเขียนเป็นสูตรการคำนวณได้ดังนี้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N \sum_{m=1}^M |R_{mnk} - \bar{R}_{nk}|}{KN(M-1)(I-1)} \dots\dots 7$$

โดยที่

RAI แทน ดัชนีความสอดคล้องระหว่างผู้ประเมิน

R_{mnk} แทน คะแนนที่ได้จากผู้ประเมินคนที่ m ของนักเรียนคนที่ n ในพฤติกรรมที่ k (m = 1, 2, 3,..., M ; n = 1, 2, 3,..., N และ k = 1, 2, 3,...,K)

\bar{R}_{nk} แทน คะแนนเฉลี่ยของนักเรียนคนที่ n ในพฤติกรรมที่ k

ซึ่งคำนวณได้จากสูตร

$$\bar{R}_{nk} = \frac{\sum_{m=1}^M R_{mnk}}{M}$$

K แทน จำนวนของพฤติกรรมบ่งชี้ทั้งหมด

N แทน จำนวนของนักเรียนทั้งหมด

M แทน จำนวนของผู้ประเมินทั้งหมด

I แทน จำนวนของคะแนนทั้งหมดที่เป็นไปได้

(ตามเกณฑ์การให้คะแนน)

ตัวอย่าง

คะแนนทักษะการพูดหน้าชั้นของนักเรียน 5 คน ที่ได้จากผู้ประเมิน 3 คน ปรากฏผลดังนี้

คนที่	คะแนนที่ได้จากผู้ประเมินคนที่ 1 (R_{1nk})				คะแนนที่ได้จากผู้ประเมินคนที่ 2 (R_{2nk})				คะแนนที่ได้จากผู้ประเมินคนที่ 3 (R_{3nk})			
	พฤติกรรมที่ 1	พฤติกรรมที่ 2	พฤติกรรมที่ 3	พฤติกรรมที่ 4	พฤติกรรมที่ 1	พฤติกรรมที่ 2	พฤติกรรมที่ 3	พฤติกรรมที่ 4	พฤติกรรมที่ 1	พฤติกรรมที่ 2	พฤติกรรมที่ 3	พฤติกรรมที่ 4
1	2	1	3	2	3	1	2	2	2	1	3	2
2	1	2	3	1	3	2	1	2	2	2	2	2
3	1	2	2	3	1	2	2	3	2	3	2	3
4	2	1	3	1	3	2	1	2	3	2	2	2
5	0	2	3	2	1	2	3	2	0	2	2	3

จากข้อมูลทั้งหมด สามารถคำนวณหาคะแนนเฉลี่ยของนักเรียนคนที่ n ในพฤติกรรมที่ k (\bar{R}_{nk}) ได้ดังนี้

นักเรียนคนที่	คะแนนเฉลี่ย (\bar{R}_{nk})			
	พฤติกรรมที่ 1	พฤติกรรมที่ 2	พฤติกรรมที่ 3	พฤติกรรมที่ 4
1	2.33	1	2.67	2
2	2	2	2	1.67
3	1.33	2.33	2	3
4	2.67	1.67	2	1.67
5	0.33	2	2.67	2.33

จากข้อมูลที่ปรากฏแสดงว่า ค่าเฉลี่ยของคะแนนที่ได้มาจากผู้ประเมินทุกคน ในพฤติกรรมที่ 1 ถึง 4 ของนักเรียนคนแรก มีค่าเท่ากับ 2.33, 1, 2.67 และ 2 ตามลำดับ สำหรับกรณีอื่น ๆ ก็มีความหมายในทำนองเดียวกัน ซึ่งโดยหลักการพื้นฐานตามที่เคยกล่าวมาแล้วนั้น หากผู้ประเมินทั้งสามคนสามารถให้คะแนนได้อย่างสอดคล้องกันมาก คะแนนเบี่ยงเบนจะต้องมีค่าเข้าใกล้ 0 เป็นส่วนใหญ่ ซึ่งจากข้อมูลที่ปรากฏกลับไม่ได้มีความสอดคล้องกันมากนัก เช่น ในการประเมินพฤติกรรมที่ 1 ของนักเรียนคนแรก ซึ่งได้คะแนนเฉลี่ยเท่ากับ 2.33 นั้น ผู้ประเมิน 3 คน ให้ 2, 3, และ 2 คะแนน คะแนนเบี่ยงเบนจึงเท่ากับ -0.33, 0.67 และ -0.33 ตามลำดับ

ในส่วนของพฤติกรรมอื่นและนักเรียนคนอื่น ๆ ก็สามารถคำนวณได้ในทำนองเดียวกัน ซึ่งเมื่อหาค่าสัมบูรณ์ของคะแนนเบี่ยงเบนและรวมค่าสัมบูรณ์ทั้งหมดจากทุกค่าหรือทุกช่อง (cell) แล้ว เราก็จะทราบได้ถึงขนาดของ

ความเบี่ยงเบนทั้งหมดที่ปรากฏจริง ซึ่งหากมีค่าสูงย่อมแสดงให้เห็นว่ามีความแตกต่างหรือมีความไม่สอดคล้องระหว่างกลุ่มผู้ประเมินนั่นเอง จากกรณีตัวอย่างจึงสามารถคำนวณหาค่าสัมบูรณ์ของคะแนนเบี่ยงเบนได้ดังนี้

n \ k	$ R_{1nk} - \bar{R}_{nk} $				$ R_{2nk} - \bar{R}_{nk} $				$ R_{3nk} - \bar{R}_{nk} $			
	1	2	3	4	1	2	3	4	1	2	3	4
1	0.33	0	0.33	0	0.67	0	0.67	0	0.33	0	0.33	0
2	1	0	1	0.67	1	0	1	0.33	0	0	0	0.33
3	0.33	0.33	0	0	0.33	0.33	0	0	0.67	0.67	0	0
4	0.67	0.67	1	0.67	0.33	0.33	1	0.33	0.33	0.33	0	0.33
5	0.33	0	0.33	0.33	0.67	0	0.33	0.33	0.33	0	0.67	0.67

$$\sum_{k=1}^K \sum_{n=1}^N \sum_{m=1}^M |R_{mnk} - \bar{R}_{nk}| = 20.63$$

แทนค่าต่าง ๆ ลงในสูตรคำนวณ จะได้

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N \sum_{m=1}^M |R_{mnk} - \bar{R}_{nk}|}{KN(M-1)(I-1)}$$

$$= 1 - \frac{20.63}{(4)(5)(3-1)(4-1)}$$

$$= 0.83$$

ดัชนีความสอดคล้องระหว่างผู้ประเมินจึงมีค่าเท่ากับ 0.83 ซึ่งแสดงให้เห็นว่าแบบวัดทักษะการพูดหน้าชั้นฉบับนี้มีคุณภาพอยู่ในขั้นใช้ได้ แต่หากปรับปรุงเกณฑ์การให้คะแนนในบางพฤติกรรมให้ชัดเจนจนสามารถสื่อความได้ตรงกันมากกว่านี้ก็จะดีมาก โดยเฉพาะพฤติกรรมที่ 1 ซึ่งมีผลรวมของค่าสัมบูรณ์ของคะแนนเบี่ยงเบนสูงที่สุด

จากสูตรคำนวณทั้งหมดที่กล่าวมาจึงเห็นได้ว่าดัชนีความสอดคล้องระหว่างผู้ประเมินไม่ได้อาศัยหลักการและขั้นตอนการคำนวณที่ยุ่งยากแต่ประการใด ผู้

เขียนจึงมีความเชื่อว่าแม้แต่ผู้ที่ไม่ค่อยชื่นชอบหรือสันทัดทางด้านคณิตศาสตร์มากนัก หากได้ลองทำความเข้าใจขั้นตอนการคำนวณดังกล่าวอย่างละเอียดสักหนึ่งหรือสองครั้งก็น่าจะช่วยให้เกิดความเข้าใจในแต่ละประเด็นได้โดยไม่ยากนัก และเนื่องจากในโลกยุคปัจจุบันเรามีเครื่องช่วยคำนวณต่าง ๆ อยู่เป็นจำนวนมากจึงเชื่ออีกว่าถ้าสามารถทำความเข้าใจหลักการพื้นฐานจนทราบถึงที่มาที่ไปของสูตรเหล่านี้ได้แล้ว การคำนวณค่าต่าง ๆ ที่เกิดขึ้นถัดจากนี้คงไม่ใช่เรื่องยุ่งยากแต่ประการใด ดังนั้นจึงเป็นสิ่งสมควรอย่างยิ่งที่ครูจะได้ใช้ความอุตสาหะของตนในการสร้างและพัฒนาเครื่องมือวัดให้มีคุณภาพสูงสุดเท่าที่พึงจะกระทำได้ ทั้งนี้เพื่อช่วยกันเพิ่มความน่าเชื่อถือให้กับกระบวนการวัดผลการเรียนรู้ ซึ่งจะส่งผลสืบเนื่องไปถึงมาตรฐานของการจัดการศึกษาโดยรวมได้ในอนาคต

เอกสารอ้างอิง

รุ่งศรีรินทร์ จันท์หอม. (2545). การสร้างเกณฑ์การประเมินผลงานทางศหกรรมศาสตร์ด้วยรูบรีค (Rubrics Assessment). วารสารศึกษาศาสตร์, 14(1), 23-29.

สุวิมล ว่องวานิช. (2535). หน่วยที่ 12 การสร้างเครื่อง
มือวัดผลด้านทักษะพิสัย. เอกสารประกอบการสอน
ชุดวิชาการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการ
เรียน (หน้า 481-558). กรุงเทพฯ : มหาวิทยาลัย
สุโขทัยธรรมาธิราช.

Burry-Stock, Judith A. and other. (1996). Rater
Agreement Indexes for Performance Assess-
ment. **Educational and Psychological
Measurement**, 56(2), 251-262.