

การวัดผลและการประเมินผล

ผศ. ดร. นภา หลิมรัตนัน*

การวัดผล (Measurement / Assessment) และการประเมินผล (Evaluation) เป็นองค์ประกอบหนึ่งของไตรยางค์การศึกษา (educational trilogy) ซึ่งประกอบด้วย 3 องค์ประกอบ คือ วัตถุประสงค์การศึกษา (educational objectives) การจัดประสบการณ์การเรียนรู้ (learning experiences) และการวัดผลและการประเมินผล และรู้จักกันโดยทั่วไปว่า OLE (O = Objectives; L = Learning experiences; E = Evaluation) ทั้ง 3 องค์ประกอบนี้มีความเกี่ยวข้องสัมพันธ์กันดังภาพต่อไปนี้



ในการจัดการเรียนการสอนโดยทั่วไป ผู้สอนจะต้องตั้งวัตถุประสงค์การศึกษาว่า สิ่งที่ต้องการสอนคืออะไร เรื่องราวดังกล่าวเป็นองค์ความรู้ (knowledge/cognitive domain) หรือเป็นทักษะ (skill / psychomotor domain) หรือเป็นทัศนคติ/เจตคติ (attitude/affective domain) เพราะวัตถุประสงค์ที่จะสอน มีส่วนสำคัญอย่างยิ่งในการกำหนดเลือกวิธีการสอน หรือ วิธีการจัดประสบการณ์การเรียนรู้แก่ผู้เรียน ตัวอย่างเช่น หากสิ่งที่ต้องการสอนเป็นเนื้อหาหรือเป็นองค์ความรู้ ผู้สอนอาจเลือกวิธีการสอนแบบบรรยาย แต่ถ้าต้องการสอนทักษะ ผู้สอนอาจเลือกสอนโดยเริ่มต้นด้วยการบรรยายเพื่อปูพื้นฐานหลักการ ข้อจำกัด ข้อควรระวังของทักษะดังกล่าว จากนั้นคงต้องมีการฝึกปฏิบัติ เพื่อให้ผู้เรียนได้มีโอกาสลงมือทำฝึกทำ ให้เกิดทักษะดังที่ต้องการ และหากต้องการสอนเจตคติหรือทัศนคติ ผู้สอนอาจต้องพิจารณารูปแบบการสอนเป็นการอภิปรายกลุ่มย่อย การถกเถียงปัญหา การเล่นเกมบทบาทสมมติ เป็นต้น ไม่ว่าวัตถุประสงค์การศึกษาจะเป็น domain ใดก็ตาม และผู้สอนคาดว่าจะได้จัดประสบการณ์การเรียนรู้แก่ผู้เรียนได้อย่างเหมาะสมตามชนิดของวัตถุประสงค์แล้ว ก็ควรจะต้องมี

* ผู้ช่วยคณบดีฝ่ายวิชาการด้านแพทยศาสตร์ศึกษา และภาควิชาชีวเคมี คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ขอนแก่น 42000

การวัดผลและประเมินผล เพื่อตรวจสอบว่าผู้เรียนได้เรียนรู้จริงตามวัตถุประสงค์หรือไม่

หากผลการประเมินไม่เป็นไปตามคาดหมาย ผู้สอนคงต้องพิจารณาว่าปัจจัยใดใน OLE ที่ส่งผลให้ผู้เรียนไม่สามารถเรียนรู้จนบรรลุวัตถุประสงค์ เป็นไปได้หรือไม่ว่า...การเขียนหรือการกำหนดวัตถุประสงค์ไม่ชัดเจน ส่งผลให้เลือกวิธีการสอนที่ไม่เหมาะสม หรืออาจเกิดจากวัตถุประสงค์ชัดเจนแต่เลือกวิธีการสอนที่ไม่สอดคล้องไม่เอื้อให้เกิดการเรียนรู้ตามวัตถุประสงค์นั้น หรืออาจเกิดจากการเลือกเครื่องมือวัดผลที่ไม่เหมาะสม ดังนั้นจะเห็นว่า ทั้ง 3 องค์ประกอบในไตรยางค์การศึกษามีส่วนเกี่ยวข้องกันอย่างแยกไม่ออกแต่ในเอกสารนี้จะขอนำเสนอรายละเอียดเฉพาะในส่วนของการวัดผลและการประเมินผลเท่านั้น โดยอยู่บนสมมติฐานว่า วัตถุประสงค์การศึกษาและการจัดประสบการณ์การเรียนรู้เหมาะสมดีแล้ว

ทุกครั้งที่ จะทำการวัดผลและประเมินผล ซึ่งมักเรียกทั่ว ๆ ไปว่า “ทำการสอบ” เราในฐานะผู้สอนจะต้องตอบคำถามต่อไปนี้เป็นอย่างน้อยเพื่อที่จะทำให้การจัดสอบนั้น ๆ เหมาะสมและยุติธรรมแก่ผู้ถูกสอบหรือผู้เรียน ในเรื่องการสอบ “ความยุติธรรม” เป็นสิ่งที่ผู้สอนพึงตระหนักอยู่เสมอ

ก่อนอื่นขอความเข้าใจเกี่ยวกับคำ (term) ในเรื่องการวัดผลและการประเมินผล 2 คำ ดังนี้คือ การวัดผล หรือ การวัด หรือ measurement หรือ assessment มีความหมายว่า การจัดหาข้อมูล หรือ จัดหาคะแนนจากหลากหลายวิธี ส่วนการประเมินผลคือการนำผลที่วัดได้มาตัดสินว่ามีคุณค่าอย่างไร เช่น ในการสอบครั้งหนึ่ง นักศึกษา ก. สอบได้คะแนน 70% ค่าคะแนน 70% คือค่าคะแนนที่วัดได้ = assessment จากนั้นผู้สอนจึงพิจารณาอีกครั้งว่าค่าคะแนน 70% นี้อยู่ในเกณฑ์ดีมาก หรือ ดี หรือ พอใช้ เราเรียกการตัดสินคุณค่าของสิ่งที่วัดได้ว่าการประเมินผล = evaluation เพราะฉะนั้นความสัมพันธ์ระหว่าง assessment และ evaluation เป็นดังนี้

Measurement = Assessment

Evaluation = Measurement (Assessment) + Judgment

อย่างไรก็ตามคำ 2 คำนี้ในบางครั้งก็สามารถใช้แทนกันได้ แต่สิ่งที่ผู้สอนต้องคำนึงถึงคือ “การวัดผลที่ดี มีโอกาสนำไปสู่การประเมินผลที่ถูกต้อง”

คำถามที่ต้องการคำตอบ

ก่อนทำการสอบผู้เรียนทุกครั้ง ผู้สอนควรตอบคำถามอย่างน้อย 5 ข้อนี้ คือ

- Why ? = ทำไมต้องสอบ ?
- What ? = สอบอะไร ?
- When ? = สอบเมื่อไหร่ ?
- Who ? = สอบใคร ?
- How ? = ใช้เครื่องมืออะไร ?

1. ทำไมต้องสอบ ?

เหตุผลที่ต้องทำการสอบผู้เรียนคือ

- ต้องการทราบว่าผู้เรียนได้เกิดการเรียนรู้ตามวัตถุประสงค์หรือไม่
- สามารถใช้ผลการสอบเป็นข้อมูลย้อนกลับ (feedback) แก่ผู้สอน เพื่อนำไปปรับปรุงกระบวนการจัดการเรียนการสอน
- สามารถใช้ผลการสอบเป็นข้อมูลย้อนกลับแก่ผู้เรียน เพื่อนำไปปรับปรุงวิธีการเรียน และเพื่อทราบความก้าวหน้าในการเรียนรู้ของตนเอง รู้จุดอ่อน-จุดแข็ง
- ใช้เป็นแรงจูงใจให้ผู้เรียนทำการศึกษา / เรียน (motivate students to study)
- ใช้จัดกลุ่มผู้เรียน (placement evaluation)
- ใช้ตัดสินผลการเรียนของผู้เรียน (ได้-ตก; A / B / C / ...)
- ใช้เป็นกลไกหนึ่งในการประกันคุณภาพของผู้เรียนต่อสังคม

ผู้สอนคงต้องตอบตัวเองว่าการจัดสอบของเราในแต่ละครั้งหวังผลในข้อใดข้างต้น ซึ่งจะเป็นตัวกำหนดถึงเนื้อหาเวลาเครื่องมือที่จะใช้ในการสอบได้อย่างเหมาะสม

2. สอบอะไร ?

สิ่งที่จะทำการสอบผู้เรียนคือ วัตถุประสงค์ของรายวิชานั้น ๆ ซึ่งได้กล่าวไปข้างแล้วในข้างต้นว่า วัตถุประสงค์ทางการศึกษา สามารถจำแนกเป็น 3 หมวดหมู่ (domain) คือ

ความรู้ = Cognitive domain = Knowledge

ทักษะ = Psychomotor domain = Skill

ทัศนคติ = Affective domain = Attitude

ผู้สอนคงต้องกำหนดให้ชัดเจนว่าต้องการวัด domain ใดจึงจะสามารถเลือกใช้เครื่องมือวัดที่เหมาะสมได้ ส่งผลให้การวัดน่าเชื่อถือและยุติธรรมต่อผู้เรียน ตัวอย่างเช่น ต้องการทดสอบว่าผู้เรียนว่ายน้ำเป็นหรือไม่ ซึ่งเป็นความสามารถทางทักษะ หากใช้เครื่องวัดเป็นการสอบข้อเขียน (paper-pencil exam) ก็คงไม่สามารถสะท้อนความเป็นจริงที่ว่าผู้ที่ตอบข้อสอบได้คะแนนดีสามารถว่ายน้ำได้จริงหรือไม่ ดังนั้นการวัดผลควรคำนึงถึงการเลือกใช้เครื่องมือที่สอดคล้องกับสิ่งที่ต้องการวัด เพื่อให้การวัดผลการประเมินผลนั้นเป็นการวัดผลตามการประเมินผลตามสภาพความเป็นจริง (authentic assessment)

3. สอบเมื่อไหร่ ?

เนื่องจากเราสามารถนำผลการสอบมาใช้เป็นข้อมูลย้อนกลับแก่ผู้เรียนให้รู้ถึงความก้าวหน้าในการเรียนและใช้ปรับปรุงวิธีการเรียน (feedback to improve) เราเรียกการสอบเพื่อวัตถุประสงค์นี้ว่า formative evaluation/diagnostic evaluation ส่วนการสอบที่หวังผลในการตัดสิน ได้-ตก หรือการให้ค่าคะแนน (grading) ซึ่งจัดเป็น final grade of record เราเรียกการสอบแบบหลังนี้ว่า Summative evaluation / Certifying evaluation เมื่อทราบถึงวัตถุประสงค์ของการสอบ 2 ชนิดข้างต้นนี้ ก็สามารถช่วยให้ผู้สอนตัดสินใจว่าจะจัดสอบในช่วงใด รวมทั้งหากต้องการสอบเพื่อจัดกลุ่มผู้เรียน (placement evaluation) ผู้สอนก็จะต้องคำนึงถึงช่วงเวลาการทดสอบเช่นกัน

หากคำนึงถึงช่วงเวลาที่ใช้จัดสอบ คงพอจะแบ่งเป็นช่วงเวลาหายบ ๆ ดังนี้คือ

- ก่อนเริ่มกระบวนการเรียนการสอนในรายวิชา (beginning of course)

- ในระหว่างที่รายวิชาดำเนินอยู่ (in course) และ

- เมื่อสิ้นสุดรายวิชา (end of course)

ดังนั้น หากเราต้องการจัดสอบเพื่อ formative evaluation คงต้องจัดสอบในระหว่างรายวิชายังดำเนินอยู่ หากจัดสอบเมื่อสิ้นสุดรายวิชาแล้วคงไม่มีประโยชน์เป็นต้น

4. สอบใคร ?

กลุ่มผู้เรียนที่ถูกสอบเป็นผู้เรียนชั้นใด กลุ่มใด ทั้งนี้เพื่อให้ผู้สอนได้คำนึงถึงความยากง่ายและความลึกซึ้งของวัตถุประสงค์ที่ทำการสอบว่าเหมาะสมหรือไม่อย่างไร ทั้งนี้เพื่อให้เกิดความยุติธรรมต่อผู้เรียน

5. ใช้เครื่องมืออะไร ?

โดยทั่วไป เครื่องมือที่ใช้ในการสอบก็คือสิ่งที่เราเรียกโดยทั่วไปว่า ข้อสอบ (test format) ซึ่งมีหลายชนิด ไม่ว่าจะเป็นชนิดใดก็ตาม เครื่องมือ/ข้อสอบ ที่ดีต้องมีคุณสมบัติดังต่อไปนี้

1. มีความเที่ยงตรง (Validity) ซึ่งแบ่งออกเป็น 4 ชนิด คือ

- Constructive validity ความเที่ยงตรงในการสร้างข้อสอบให้สามารถวัดพฤติกรรมการเรียนรู้ของผู้เรียนได้ บางครั้งเรียกว่า ความเที่ยงตรงตามทฤษฎี

Constructive validity ของเครื่องมือชนิดต่าง ๆ

	Triple jump	MEQ	Direct Observation	MCQ	OSCE
- Knowledge	+	++		+++	+
- Problem Solving	+++	++	+	+	+
- Interpersonal skill			+++		+
- Technical skill			++		+++
- Attitudes	±		+		

- Content validity ความเที่ยงตรงตามเนื้อหา หมายถึงมีการสุ่มเนื้อหาที่ออกสอบได้เหมาะสมจนสามารถเป็นตัวแทนของรายวิชานั้น ๆ (adequate sampling of content) เราสามารถใช้ตารางเนื้อหา (test blueprint/table of specification) ช่วยในการให้ได้มาซึ่งข้อสอบที่มี content validity โดยทั่วไปมีข้อเสนอแนะว่า ในชุดข้อสอบใดๆ ควรประกอบด้วยเนื้อหาที่

ต้องรู้ ~ 60-85%

ควรรู้ ~ 10-35%

น่าจะรู้ ~ 5-10%

- Concurrent validity ความเที่ยงตรงตามสถานการณ์ เช่น ข้อสอบนี้วัดหลักการการใช้เครื่องมือ A ดังนั้นหากนักศึกษาที่ได้คะแนนสูง ควรใช้เครื่องมือ A เป็น และในทางกลับกัน นักศึกษาที่ได้คะแนนต่ำควรใช้เครื่องมือ A ไม่เป็น หรือไม่ถูกต้อง

- Predictive validity ความเที่ยงตรงในการคาดเดา เช่น หากข้อสอบชุดนี้มี predictive validity ดี หมายความว่านักศึกษาที่สอบได้คะแนนสูง คงไม่มีปัญหาในการเรียน และนักศึกษาที่ได้คะแนนต่ำ อาจมีปัญหาในการเรียน เป็นต้น

2. มีความเป็นปรนัย (Objectivity) หมายถึง ผลของการสอบชุดข้อสอบนั้น ๆ สะท้อนถึงความสามารถของผู้เรียนอย่างแท้จริง ไม่ได้มีอิทธิพลของผู้สอนเข้ามาเกี่ยวข้อง อิทธิพลอันเนื่องมาจากผู้สอนเรียกว่า subjectivity “ข้อสอบที่ดีควรมี objectivity สูงๆ และมี subjectivity ต่ำๆ”

ความเป็นปรนัย ได้แก่

- ความเป็นปรนัยในการถาม หรือ ความชัดเจนในการถาม คือ อ่านแล้วเข้าใจตรงกัน ไม่ต้องการการตีความเพิ่มเติม

- ความเป็นปรนัยในการให้คะแนน หรือ ความชัดเจนในการให้คะแนน หมายถึง ตรวจแล้วให้คะแนนตรงกัน ไม่ว่าผู้ตรวจจะเป็นใคร เท่ากับเป็นการสะท้อนผลการกระทำของผู้เรียนเท่านั้น ตัวอย่างชนิดของข้อ

สอบที่มีความเป็นปรนัยสูงในการให้คะแนน คือ ข้อสอบแบบเลือกตอบ (Multiple choice question: MCQ) ส่วนข้อสอบที่มีความเป็นปรนัยต่ำในการให้คะแนน ได้แก่ ข้อสอบแบบบรรยาย (essay) เป็นต้น

ความเป็นปรนัย (objectivity) และความเป็นอัตนัย (subjectivity) เป็นคุณสมบัติของข้อสอบ ไม่ใช่รูปแบบข้อสอบแบบเลือกตอบ (MCQ) หากไม่ระมัดระวังในการออก ก็อาจกลายเป็นข้อสอบที่ไม่มีความเป็นปรนัยก็ได้และในทางตรงกันข้ามข้อสอบแบบบรรยาย (essay) หากระมัดระวังในการออกโดยใช้คำพูดที่ชัดเจนจำเพาะเจาะจง ก็อาจเป็นข้อสอบที่มีความเป็นปรนัยได้เช่นกัน ความเป็น “ปรนัย/ อัตนัย” อาจทำให้ผู้สอนมีความเข้าใจคลาดเคลื่อน เนื่องจากเราได้มีการกำหนดเรียกข้อสอบชนิดเลือกตอบ (MCQ) ว่าข้อสอบปรนัย และเรียกข้อสอบแบบบรรยาย (essay) ว่าข้อสอบอัตนัย โดยนำคำเหล่านั้นผูกเข้ากับรูปแบบ (test format) ทั้งๆ ที่ในความเป็นจริงเป็นคุณสมบัติของข้อสอบ

3. มีความเชื่อถือได้ (Reliability) หมายความว่า หากนำข้อสอบชุดเดิมมาทำการสอบ 2 ครั้ง แล้วให้ผลเหมือนเดิม ซึ่งเป็นการยากที่จะทำการทดสอบชุดข้อสอบใดๆ 2 ครั้ง โดยผู้สอบกลุ่มเดิมในช่วงเวลาห่างกันพอสมควร (1-2 สัปดาห์) โดยผู้สอบไม่รู้ล่วงหน้า ทั้งนี้ปัจจัยต่าง ๆ เกี่ยวกับผู้สอบจะต้องเหมือนเดิมทุกอย่าง การทดสอบความเชื่อถือ/เชื่อมั่นของข้อสอบแบบนี้เรียกว่าการทำ Test-Retest ดังนั้น นักการศึกษาจึงได้คิดวิธีสะท้อนค่าความเชื่อถือของชุดข้อสอบ โดยใช้ค่าสถิติในการหาค่า reliability โดยไม่ต้องให้ผู้สอบสอบ 2 ครั้งเรียกว่าค่าความเชื่อมั่นดังกล่าวว่า ค่าคงที่ภายใน (internal consistency) ใช้สัญลักษณ์ว่า α สูตรในการหาค่า α มีหลายสูตร เช่น KR 20, KR 21, alpha coefficient (Cronbach alpha method) เป็นต้น

4. มีความยากง่ายพอเหมาะ (Difficulty index) โดยมีข้อเสนอแนะว่าในแต่ละชุดข้อสอบควรมีข้อสอบที่มีค่าความยากง่ายปานกลาง ~ 50%; ยาก ~ 25%; และง่าย ~ 25%

5. มีค่าอำนาจจำแนกดี (Discrimination power)

หมายถึง สามารถจำแนกคนเก่ง-อ่อน ออกจากกันได้ รายละเอียดเกี่ยวกับการหาค่าง่ายของข้อสอบและค่าอำนาจจำแนก ผู้สนใจสามารถศึกษาได้ในหัวข้อการวิเคราะห์ข้อสอบ (Item analysis) ในหนังสือเกี่ยวกับการวัดผลการศึกษาโดยทั่วไป

6. สร้างง่าย ใช้ง่าย ใช้สะดวก คุ่มค่า (Feasibility, Practicability, Usability)

ชนิดของข้อสอบ (Test format)

แบ่งเป็น 2 จำพวกใหญ่คือ Limited choice item (question) และ Open-ended item (question)

- Limited choice Item ได้แก่ multiple choice question, true/false question, matching column เป็นต้น

- Open-ended item ได้แก่ sentence completion, short answer question (SAQ), modified essay question (MEQ), และ essay เป็นต้น

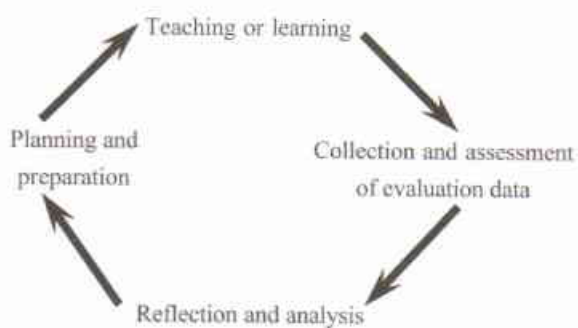
ข้อสอบสองจำพวกนี้มีข้อดี-ข้อเสียแตกต่างกันดังต่อไปนี้

Open-ended Item	Limited Choice Item
- สามารถวัดความสามารถของผู้เรียนในระดับที่สูงกว่าความจำ เช่น ความเข้าใจ, ทักษะการแก้ปัญหา เป็นต้น	- วัดความสามารถในระดับความจำเป็นส่วนใหญ่
- ครอบคลุมเนื้อหา (content coverage) ได้ค่อนข้างจำกัด	- ครอบคลุมเนื้อหาได้มากกว่า
- ตรวจ-ให้คะแนนยาก (ความเป็นปรนัยในการให้คะแนนค่อนข้างต่ำ)	- ตรวจ-ให้คะแนนง่าย
- สามารถได้ข้อมูลย้อนกลับที่ชัดเจน	- ข้อมูลย้อนกลับที่ได้คลุมเครือ ไม่ชัดเจน
- ป้องกันการทุจริตในการสอบได้ดีกว่า	- ยากต่อการทุจริตในการสอบ

สรุป

การวัดผลการประเมินผล จัดเป็นองค์ประกอบสำคัญอย่างยิ่งในไตรยางค์การศึกษา จนกระทั่งมีการกล่าวอ้างว่า “Evaluations shape up study” ผู้เรียนจะเรียนตามสิ่งที่เขาคาดว่าจะถูกสอบ ดังนั้นหากผู้สอนระมัดระวังหรือให้ความสำคัญในการออกข้อสอบ ก็สามารถกำกับการเรียนรู้ไปในทิศทางที่ต้องการได้ด้วยนอกเหนือจากการจัดประสบการณ์การเรียนรู้ และคุณจะมีประสิทธิภาพในการกำกับการเรียนรู้ของผู้เรียนมากกว่าด้วยซ้ำ สิ่งที่คุณสอนควรจะต้องตระหนักเกี่ยวกับการประเมินผลคือ

1. ต้องชัดเจนว่าต้องการวัดอะไร เพื่ออะไร
2. ต้องเลือกเครื่องมือที่ใช้ในการวัดผลให้เหมาะสม สอดคล้องกับสิ่งที่ต้องการวัด
3. ควรใช้เครื่องมือที่หลากหลาย เพราะเครื่องมือแต่ละชนิดมีข้อจำกัดแตกต่างกัน
4. ตระหนักว่าการวัดผลไม่ใช่จุดสิ้นสุดของการเรียนการสอน ควรนำผลไปปรับปรุงเปลี่ยนแปลงการเรียนการสอนในโอกาสต่อไปอย่างต่อเนื่องดังวงจรการวัดผล (evaluation cycle) ต่อไปนี้



ผู้ที่สามารถบอกว่าข้อสอบดีหรือไม่ ได้แก่ ผู้รู้ในเนื้อหานั้น ๆ (content expert) และนักการศึกษาที่มีความรู้เรื่องการวัดผล (test technician) ผู้สอนที่มีได้ศึกษาเกี่ยวกับการศึกษาโดยตรงหากทำความเข้าใจ/เรียนรู้เกี่ยวกับเรื่องเหล่านี้ก็สามารถมีบทบาทเป็น test technician ไปในตัวได้ด้วย ทั้งนี้เพื่อส่งผลให้ได้ข้อสอบที่ดีได้ผลการวัดที่น่าเชื่อถือ และนำไปสู่การประเมินผลที่ถูกต้อง

เอกสารประกอบการเรียนเรียง

1. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination. *Med Educ* 1979; 13:41-54.
2. เอกสารประกอบการอบรมเชิงปฏิบัติการเรื่อง Assessing Aspects of Competence: A course for Medical and Paramedical Teacher; March 18-22, 1991; University of Dundee, Scotland. 1991.
3. เอกสารประกอบการอบรมเชิงปฏิบัติการเรื่อง Evaluation of student performance in medical and health profession education; Mar 29-Apr 2, 1993; Ben-Gurion University, Beersheba, Israel. 1993.
4. McGuire CH. Reflections of a maverick measurement maven. *JAMA* 1995; 274: 735-40.
5. Wilkes M, Bligh J. Evaluating educational intervention. *BMJ* 1999; 318:1269-72.
6. Available from: http://www.bamaed.ua.edu/ail601/student_evaluation.htm
7. Available from: <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as81k9.htm>
8. Available from: <http://ericacve.org/docs/auth-pab.htm>