Canadian Medical Education Journal

Black Ice

Five ways to get a grip on grouped self-assessments of competence for program evaluation

Cinq façons pour mieux comprendre les auto-évaluations groupées des compétences pour l'évaluation de programme

Rebecca Zhao,¹ Marcel D'Eon¹

¹University of Saskatchewan

Published ahead of issue: March 9, 2020; published: August 6, 2020

CMEJ 2020, 11(4), e90-e96, Available at http://www.cmej.ca

© 2020 Zhao, D'Eon; licensee Synergies Partners

https://doi.org/10.36834/cmej.69276

This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Self-assessments conducted by individuals when taken together (grouped) provide valid and accurate measures of learning outcomes of the group. This is useful for program evaluation. Grouped self-assessments are simple to understand and construct, easy to implement, relatively accurate, and do not require extensive and complex prepost testing measures. However, group self-assessments have the potential to be misused. To examine how group self-assessments have been used in medical education, we conducted a search of journal articles published in 2017 and 2018 from eight prominent medical education journals. Twenty-seven (n = 27) articles that used self-assessments for program evaluation were selected for data extraction and analysis. We found three main areas where misuse of self-assessments may have resulted in inaccurate measures of learning outcomes: measures of "confidence" or "comfort", pre-post self-assessments, and the use of ambiguous learning objectives. To prevent future misuse and to build towards more valid and reliable data for program evaluations, we present the following recommendations: measure competence instead of confidence or comfort; use pre-test self-assessments for instructional purposes only (and not for data); ask participants to do the post-intervention self-assessments first followed by retrospective pre-intervention self-assessments afterwards; and use observable, clear, specific learning objectives in the educational intervention that can then be used to create the self-assessment statements.

Résumé

Les auto-évaluations effectuées individuellement, quand mises en commun(regroupées) fournissent des mesures valides et précises des résultats de l'apprentissage du groupe. Ceci s'avère utile dans le cadre d'évaluation de programme. Les auto-évaluations groupées sont faciles à comprendre, à créer et à mettre en œuvre. De plus, elles sont relativement précises et ne requièrent pas de mesures exhaustives et complexes avant et après l'évaluation.

Cependant, il est possible de mal utiliser les auto-évaluations de groupe. Afin d'examiner de quelle façon les autoévaluations de groupe ont été utilisées dans la formation médicale, nous avons dépouillé des articles de huit revues médicales importantes publiées en 2017 et 2018. Vingt-sept (n=27) articles ayant utilisé les auto-évaluations à des fins d'évaluation de programme ont été sélectionnés pour être analysées. Nous avons trouvé trois domaines principaux où la mauvaise utilisation des auto-évaluations a peut-être occasionné des mesures inexactes des résultats de l'apprentissage : mesures de « confiance » ou d'« aisance », auto-évaluations pré/post et l'emploi d'objectifs d'apprentissage ambigus. Afin de prévenir l'utilisation future incorrecte et de miser sur des données plus valides et fiables à des fins d'évaluation de programme, nous présentons les recommandations suivantes : mesurer la compétence au lieu de la confiance ou l'aisance; utiliser les auto-évaluations prétest à des fins d'enseignement seulement (et non comme données); demander aux participants de faire les auto-évaluations post-intervention d'abord suivies par les auto-évaluations préintervention rétrospectives après; utiliser des objectifs d'apprentissage observables, clairs et spécifiques lors de l'intervention de formation qui peuvent servir de base aux items d'autoévaluation.

Grouped self-assessments (GS-a) are accurate and valid measures of learning outcomes for the target group in the context of program evaluation.¹ We use the definition of program evaluation by the Canadian Evaluation Society found on their website (https://evaluationcanada.ca/what-is-evaluation): evaluation is the systematic assessment of the design, implementation or results of an initiative for the purposes of learning or decision-making. GS-a, therefore, systematically collect data that can help determine some of the results of a learning initiative. By self-assessment, we mean the rating of one's own ability to meet a learning objective or to performs a task. For example, people may be asked to rate the extent to which they are able to describe the progression of Type 2 diabetes in a middle-aged patient. These individual self-assessments do not predict individual achievement but when grouped give a valid and reliable measure of how much the group as a whole learned from the intervention.¹GSa is the arithmetic mean of all the self-assessments completed by members of the target audience, usually a group of learners. Furthermore, GS-a are simple to construct (they are based on the objectives of the educational intervention) and easy to administer (both the pre- and post-intervention selfassessments are collected after the intervention).² However, these data collection methods could be and are misused in program evaluations. In this article, we describe how some researchers use and misuse GS-a in health professions education and explain why some of these may yield poor data for program evaluation.

To provide empirical evidence for the uses and especially misuses of self-assessments, we conducted

a quick search for articles published in 2017 and 2018 in eight medical education journals: Academic Medicine, Advances in Health Sciences Education, BMC Medical Education, Canadian Medical Education Journal, Medical Education, Medical Teacher, and Perspectives on Medical Education. Twenty-seven articles that used GS-a for program evaluation were selected for data extraction and analysis. We tallied the number of times that these features of selfassessment were used: use of grouped pre selfassessment as data; measures of comfort or confidence, competence, and readiness; comparisons of pre-test and post-test data; comparisons of retrospective pre-test and post-test data, analyses of post-test data only; and the presence of observable, clear, specific learning objectives. See Table 1.

Some studies used a topic or domain area rather than clearly written and observable learning objectives connected to the intervention. Specific observable statements of learning outcomes will give more accurate data than broad, vague statements that might be open to great variation in interpretation.

Many studies we found used "confidence" and "competence" interchangeably. Confidence suggests one's willingness to undertake an activity and to continue undertaking the task if initially unsuccessful, whereas competence alludes to what is known about one's own ability and one's previous experience of carrying out a task.³ They asked questions such as "how comfortable are you...?" Of the 27 selected reports, 18 studies measured perceived confidence in or comfort with performing certain tasks.^{4–21} Six studies measured competency or ability,^{10,13,22–25} two studies measured readiness,^{9,16} and five studies lacked sufficient detail to determine what they used.^{26–30} Measuring one's willingness to perform a task or one's confidence helps promote reflection of

performance³ but is not the most effective method of evaluating learning or performance.

Authors (Year)	Used pre-Self- assessment as data	Measured confidence and/or comfort	Measured competency	Measured readiness	Pre- vs true post-test comparison	Retrospective pre- vs post- test comparison	Post-test only	Observable, clear, specific learning objectives
Total: 27	19	18	6	2	19	4	5	22
Bartlett, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
Bartman, et a. (2018)	No	No	Yes	No	No	Yes	No	Yes
Chokshi, et al. (2017)	No	No	No	No	No	Yes	No	No
Clay, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
Clementz, et al. (2017)	No	Yes	No	No	No	No	Yes	Yes
Držaić, et al. (2018)	No	No	Yes	No	No	No	Yes	Yes
Garibaldi, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
Gomes, et al. (2017)	No	No	No	No	No	No	Yes	Yes
Kaminetzky, et al. (2017)	No	Yes	Yes	No	No	No	Yes	Yes
Lévesque, et al. (2018)	Yes	No	Yes	No	Yes	No	No	Yes
Levine, et al. (2018)	Yes	Yes	No	No	Yes	No	No	Yes
Loke, et al. (2017)	No	Yes	No	No	No	Yes	No	No
Ludwig, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
O'Donoghue, et al. (2018)	Yes	Yes	No	No	No	No	Yes	Yes
Peluso, et al. (2017)	Yes	Yes	No	Yes	Yes	No	No	Yes
Pettignano, et al. (2017)	Yes	No	No	No	Yes	No	No	No
Phillips, et al. (2017)	Yes	Yes	No	Yes	Yes	No	No	Yes
Rassbach, et al. (2018)	Yes	Yes	No	No	Yes	Yes	No	No
Richardson, et al. (2018)	Yes	Yes	No	No	Yes	No	No	Yes
Rusiecki, et al. (2018)	Yes	Yes	No	No	Yes	No	No	Yes
Sabouni, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
Shiels, et al. (2017)	No	Yes	Yes	No	Yes	No	No	Yes
Sopoaga, et al. (2017)	Yes	No	No	No	Yes	No	No	Yes
Tchekmedyian, et al. (2017)	Yes	No	Yes	No	Yes	No	No	Yes
Wilkes, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes
Yang, et al. (2017)	Yes	No	No	No	Yes	No	No	No
Yeung, et al. (2017)	Yes	Yes	No	No	Yes	No	No	Yes

Table 1. Summary of features found in grouped self-assessment studies.

Additionally, pre-test GS-a coupled with post-test GSa are inadequate measurements of program results.^{31,32} Pre-post measures are prone to response shift bias in subjects' estimates of training effects.³¹ Specifically, an intervention may change one's standard of evaluation because of changes in the participant's internal standards, values, or understandings.³³ Thus, changes in a respondent's self-evaluation standard would be reflected in their post-test self-assessment. According to our analysis, five studies had not used any pre-intervention assessments at all.^{5,10,18,22,28} Eighteen studies self-assessments,^{4,6–9,12–} pre-post administered $^{17,19,21,23,25,27,29,30}\xspace$ whereas only three studies employed retrospective pre-test selfpost-intervention assessments.^{11,24,26} One study had employed both pre-intervention pre-test and post-intervention retrospective pre-test self-assessments.²⁰ The majority of selected studies in our analyses may have been affected by the response shift bias with prepost self-assessments, suggesting inaccurate data for program evaluation.

Learning is an important purpose of all educational interventions. Thus, one should be referring to learning objectives to analyze or judge educational interventions. Some studies used a topic or domain area rather than clearly written and observable learning objectives tied to the intervention. Specific observable statements of learning outcomes will give more accurate data than broad, vague statements that might be open to great variation in interpretation. Of the 27 selected studies, 22 studies used the observable, clearly written, specific learning objectives from the program or intervention in their self-assessments.^{4–10,12–18,21–25,28,29} Yet, five studies either lacked sufficient details to determine what they used or used topic areas and broad domains instead of learning objectives.11,20,26,27,30 When selfassessments refer to ambiguous learning objectives, respondents may not be able to accurately interpret what they are expected to assess. If ambiguous learning objectives or broad topic areas are used, respondents' interpretations of what they were supposed to self-assess may differ from the interpretations of other respondents, the educators, and researchers.³⁴

How to Use Grouped Self-Assessments in Program Evaluation

- 1. Use competence as a criterion for measurement; not confidence or comfort. There is a difference between measuring competence versus confidence or comfort. The former refers to perceived ability to do an activity (as defined by the objectives of the educational initiative), whereas the latter refers to one's judgement of their willingness to do an activity. Thus, we should ask people to assess their own competence or ability to do a task and not their confidence or comfort in doing that task. We might ask, then, to what extent the learners are able to accomplish a task or can describe a phenomenon not how comfortable or confident they are in performing the task or describing the phenomenon
- 2. Have observable, clearly written, specific learning objectives in the self-assessments. When creating self-assessments, one should avoid ambiguous questions so that all respondents may interpret questions similarly. Additionally, self-assessment questions or prompts should be specific enough that respondents know what their response should be about, and researchers know what it means.³⁵
- 3. Use pre-test self-assessments for instructional purposes only. Asking learners to assess themselves before the educational intervention alerts them to the goals of instruction and prepares them for learning. This can improve the effectiveness of the intervention but not the evaluation. Subject to the response shift bias explained above, pre-test self-assessments alone or coupled with post-test self-assessments will yield inaccurate data. Clearly, the standards people use to self-assess change as they learn.
- 4. Use post-intervention retrospective pre-test self-assessments with post-intervention self-assessments. A pre-test gathers data before the intervention. A retrospective pre-test gathers data from before the intervention ("pre") but by looking back (retrospectively) after the intervention. This type of assessment will minimize the risk of the response shift bias in influencing the accuracy of the self-assessments. In the course of the educational intervention, a respondent will change the standard against which they would assess themselves. Once they

learn more, their standard goes up. Hence, as explained earlier, it is better to ask participants <u>after the intervention</u>, the extent to which they <u>were</u> able to describe the progression of diabetes in a middle-aged man <u>before the session</u> took place.

5. Following the educational intervention, ask first current condition about their (postintervention) and then ask respondents to look back to just before they started their educational sessions.³² It is best to ask for the entire set of post-intervention self-assessments first followed by the retrospective pre-test selfassessments. Do not ask for self-assessments, both pre- and post- objective by objective, as respondents may complete the questions to obtain the result according to how much they believe they have learned. We provide an example in Figure 1.

To avoid the dangerous black ice of grouped selfassessments, use our five guidelines to get a grip on your next program evaluation We have created an example self-assessment for the end of an educational session on creating logic models for program evaluation. See Figure 1. Finally, we hope that researchers will further explore Gs-a in the search for other "better" practices and for the mechanisms that make them work.

Figure 1. An example of a post-test self-assessment and a retrospective pre-test self-assessment at the end of an educational session on creating logic models for program evaluation.

To what extent are you <u>now</u> able to meet the following objectives? (1 is "not at all" and 5 is "completely")

1. Describe the important steps in creating logic models.

- 1 2 3 4 5
- 2. Critique a simple logic model. 1 2 3 4 5

To what extent <u>were</u> you able, before completing this workshop, to meet the following objectives? (1 is "not at all" and 5 is "Completely")

Describe the important steps in creating logic models.
2 3 4 5
Critique a simple logic model.
1 2 3 4 5

Conflicts of interest: The authors declare no conflicts of interest.

References

- D'Eon MF, Trinder K. Evidence for the validity of grouped self-assessments in measuring the outcomes of educational programs. *Eval Health Prof.* 2014 Dec;37(4):457-69. https://doi.org/10.1177/0163278713475868
- Schiekirka S, Reinhardt D, Beibarth T, Anders S, Pukrop T, Raupach T. Estimating learning outcomes from preand posttest student self-assessments: a longitudinal study. Acad Med. 2013 Mar;88(3):369-75. https://doi.org/10.1097/ACM.0b013e318280a6f6
- Stewart J, O'Halloran C, Barton JR, Singleton SJ, Harrigan P, Spencer J. Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. Med Educ. 2000 Nov 21;34(11):903-9. <u>https://doi.org/10.1046/j.1365-2923.2000.00728.x</u>
- Richardson KM, Singh J, Muñoz D, Damp JB, Mendes LA. Improving practice guideline adherence through peer feedback: impact of an ambulatory cardiology curriculum. *Teach Learn Med*. 2018 Jul 3;30(3):328-36. <u>https://doi.org/10.1080/10401334.2017.1415811</u>
- O'Donoghue D, Davison G, Hanna L-J, McNaughten B, Stevenson M, Thompson A. Calibration of confidence and assessed clinical skills competence in undergraduate paediatric OSCE scenarios: a mixed methods study. BMC Med Educ 2018 Dec; 18(1). https://doi.org/10.1186/s12909-018-1318-8
- Ludwig AB, Raff AC, Lin J, Schoenbaum E. Group observed structured encounter (GOSCE) for third-year medical students improves self-assessment of clinical communication. *Med Teach*. 2017 May 27;39(9):1-5. <u>https://doi.org/10.1080/0142159X.2017.1332361</u>
- Garibaldi BT, Niessen T, Gelber AC, Clark B, Lee Y, Madrazo JA, et al. A novel bedside cardiopulmonary physical diagnosis curriculum for internal medicine postgraduate training. *BMC Med Educ*. 2017 Dec 17(1). <u>https://doi.org/10.1186/s12909-017-1020-2</u>
- Bartlett RD, Radenkovic D, Mitrasinovic S, Cole A, Pavkovic I, Denn PCP, et al. A pilot study to assess the utility of a freely downloadable mobile application simulator for undergraduate clinical skills training: a single-blinded, randomised controlled trial. *BMC Med Educ* 17(1). <u>https://doi.org/10.1186/s12909-017-1085-y</u>
- 9. Peluso MJ, Luckett R, Mantzor S, Bedada AG, Saleeb P, Haverkamp M, et al. Strengthening medical training programmes by focusing on professional transitions: a

national bridging programme to prepare medical school graduates for their role as medical interns in Botswana. *BMC Med Educ* 2017 Dec;17(1). https://doi.org/10.1186/s12909-017-1102-1

- Kaminetzky CP, Beste LA, Poppe AP, Doan DB, Mun HK, Woods NF. Implementation of a novel population panel management curriculum among interprofessional health care trainees. BMC Med Educ 2017 Dec; 17(1). <u>https://doi.org/10.1186/s12909-017-1093-y</u>
- 11. Loke Y-H, Harahsheh AS, Krieger A, Olivieri LJ. Usage of 3D models of tetralogy of Fallot for medical education: impact on learning congenital heart disease. *BMC Med Educ.* 2017 Dec; 17(1). https://doi.org/10.1186/s12909-017-0889-0
- Sabouni A, Bdaiwi Y, Janoudi SL, Namous LO, Turk T, Alkhatib M, et al. Multiple strategy peer-taught evidence-based medicine course in a poor resource setting. BMC Med Educ. 2017 Dec ;17(1). https://doi.org/10.1186/s12909-017-0924-1
- Shiels L, Majmundar P, Zywot A, Sobotka J, Lau CSM, Jalonen TO. Medical student attitudes and educational interventions to prevent neurophobia: a longitudinal study. *BMC Med Educ.* 2017 Dec; 17(1). <u>https://doi.org/10.1186/s12909-017-1055-4</u>
- Wilkes MS, Day FC, Fancher TL, McDermott H, Lehman E, Bell RA, et al. Increasing confidence and changing behaviors in primary care providers engaged in genetic counselling. *BMC Med Educ*. 2017 Dec; 17(1). https://doi.org/10.1186/s12909-017-0982-4
- Yeung C, Friesen F, Farr S, Law M, Albert L. Development and implementation of a longitudinal students as teachers program: participant satisfaction and implications for medical student teaching and learning. BMC Med Educ. 2017 Dec; 17(1) https://doi.org/10.1186/s12909-017-0857-8
- Phillips AC, Mackintosh SF, Bell A, Johnston KN. Developing physiotherapy student safety skills in readiness for clinical placement using standardised patients compared with peer-role play: a pilot nonrandomised controlled trial. *BMC Med Educ.* 2017 Dec; 17(1) <u>https://doi.org/10.1186/s12909-017-0973-</u> <u>5</u>
- Clay AS, Ming DY, Knudsen NW, Engle DL, Grochowski CO, Andolsek KM, et al. CaPOW! Using problem sets in a capstone course to improve fourth-year medical students' confidence in self-directed learning: *Acad Med.* 2017 Mar;92(3):380-4. https://doi.org/10.1097/ACM.000000000001193
- Clementz L, McNamara M, Burt NM, Sparks M, Singh MK. Starting With Lucy: Focusing on human

similarities rather than differences to address health care disparities. *Acad Med.* 2017 Sep;92(9):1259-63. https://doi.org/10.1097/ACM.000000000001631

- Levine SA, Chao SH, Caruso LB, Jackson AH, Russell ML, Young ME, et al. Chief resident immersion training in the care of older adults: a successful national replication of an interspecialty educational intervention. Acad Med. 2018 Sep;93(9):1341-7. https://doi.org/10.1097/ACM.00000000002311
- Rassbach CE, Blankenburg R. A novel pediatric residency coaching program: outcomes after one year. Acad Med. 2018 Mar;93(3):430-4. https://doi.org/10.1097/ACM.000000000001825
- Rusiecki J, Schell J, Rothenberger S, Merriam S, McNeil M, Spagnoletti C. An innovative shared decisionmaking curriculum for internal medicine residents: findings from the University of Pittsburgh Medical Center. Acad Med. 2018 Jun;93(6):937-42. https://doi.org/10.1097/ACM.000000000001967
- 22. Držaić M, Kummer I, Mucalo I, Bruno A, Ortner Hadžiabdić M. Identifying self-assessed competencies and areas for improvement within community pharmacist-preceptors support during preregistration training. BMC Med Educ. 2018 Dec; 18(1). https://doi.org/10.1186/s12909-018-1413-x
- Lévesque M, Gauthier-Boudreau J, Gagnon P, Bertulies-Esposito B, Hatcher S, Gagnon L. Evaluation of an evidence-based medicine educational intervention in a regional medical campus. *Can Med Educ J.* 2018;9(1):e51-8. https://doi.org/10.36834/cmej.42225
- Bartman T, Heiser K, Bethune A, Crandall W, McClead R, Davis JT, et al. Interprofessional QI training enhances competency and QI productivity among graduates: findings from Nationwide Children's Hospital. Acad Med. 2018 Feb;93(2):292-8. https://doi.org/10.1097/ACM.000000000001862
- Tchekmedyian V, Shields HM, Pelletier SR, Pazo VC. The effect of rubric-guided, focused, personalized coaching sessions and video-recorded presentations on teaching skills among fourth-year medical students: a pilot study. *Acad Med.* 2017 Nov;92(11):1583-9. https://doi.org/10.1097/ACM.00000000001686
- Chokshi BD, Schumacher HK, Reese K, Bhansali P, Kern JR, Simmens SJ, et al. A "resident-as-teacher" curriculum using a flipped classroom approach: can a model designed for efficiency also be effective? Acad Med. 2017 Apr;92(4):511-4. https://doi.org/10.1097/ACM.00000000001534

- Pettignano R, Bliss L, McLaren S, Caley S. Interprofessional medical-legal education of medical students: assessing the benefits for addressing social determinants of health. *Acad Med.* 2017 Sep;92(9):1254-8. <u>https://doi.org/10.1097/ACM.000000000001581</u>
- Gomes AW, Butera G, Chretien KC, Kind T. The development and impact of a social media and professionalism course for medical students. *Teach Learn Med.* 2017 Jul 3;29(3):296-303. <u>https://doi.org/10.1080/10401334.2016.1275971</u>
- Sopoaga F, Zaharic T, Kokaua J, Covello S. Training a medical workforce to meet the needs of diverse minority communities. *BMC Med Educ.* 2017 Dec; 17(1) <u>https://doi.org/10.1186/s12909-017-0858-7</u>
- 30. Yang L-Y, Huang C-C, Hsu H-C, Yang Y-Y, Chang C-C, Chuang C-L, et al. Voluntary attendance of smallgroup brainstorming tutoring courses intensify new clerk's "excellence in clinical care": a pilot study. BMC Med Educ. 2017 Dec; 17(1) https://doi.org/10.1186/s12909-016-0843-6
- 31. Howard GS, Dailey PR. Response-Shift Bias: a source of contamination of self-report measures. J Appl Psychol.

1979;64(2):144-50. <u>https://doi.org/10.1037/0021-</u> 9010.64.2.144

- Nimon K, Zigarmi D, Allen J. Measures of program effectiveness based on retrospective pretest data: are all created equal? *Am J Eval.* 2011 Mar;32(1):8-28. <u>https://doi.org/10.1177/1098214010378354</u>
- Nimon K. Explaining differences between retrospective and traditional pretest selfassessments: competing theories and empirical evidence. Int J Res Method Educ. 2014 Jul 3;37(3):256-69. https://doi.org/10.1080/1743727X.2013.820644
- Eberly Center for Teaching Excellence. The educational value of course-level learning objectives/outcomes [Internet]. Carnegie Mellon University. Available from: https://www.cmu.edu/teaching/resources/Teaching/ CourseDesign/Objectives/CourseLearningObjectivesV alue.pdf [Accessed on 2019 Jun 26].
- 35. Price PC, Jhangiani R, Chang I-CA. Survey research: Constructing survey questionnaires. *Research methods in psychology*. 2nd Canadian edition. 2015.