Consistencies of Student Self-Assessment on Professional Behavior Skills Following Simulation

Authors:
Ofra Pottorf, PT, PhD, DPT, OCS[1]
Director of Clinical Education / Clinical Assistant Professor

Jason Ward, EdD[2]
Faculty of Professional Studies

Yasser Salem, PT, PhD, NCS, PCS[1]

Affiliations and Work Attributions:
[1]Hofstra University School of Health Professions and Human Services, Department of Allied Health and Kinesiology.

[2]Capella University, School of Public Service and Education

Corresponding Author:
Ofra Pottorf, PT, PhD, DPT, OCS
Ofra.pottorf@hofstra.edu
C: 516-380-9646
O: 516-463-5188
222C Hagedorn Hall
119 Hofstra University, Hempstead, NY 11549

Sponsoring Information:
This research has not been sponsored, funded, or supported by any organization
Conflict of Interests: The authors have no conflict of interest to report.
There are no previous presentations of this manuscript
Ethical Approval: IRB approval #s: 2021-87 and 2115.

Abstract

The purpose of this study was to investigate if physical therapy students' self-assessment on professional behaviors skills were consistent with faculty assessment after a simulation experience. A standard self-assessment tool on professional behaviors was completed after a simulation experience, where consistency was determined by comparison of faculty ratings. The M ± SD (SEM) was 15.84 ± 4.15 (0.83) for student ratings, and 14.60 ± 3.61 (0.72) for faculty ratings with p = .004, and Pearson's Correlation Coefficient (*r)* = 0.884. This study found strong positive correlations between student and faculty reported ratings. The professional behaviors with highest consistencies were communication, commitment to learning, stress management, and critical thinking. Those with lowest consistencies were problem solving and effective use of time and resources. Structured simulated experiences may play a role of developing self-assessment skills on professional behaviors for physical therapy students.

Key words: Self-Assessment, Consistency, Professional Behaviors, Simulation, Students.

1. INTRODUCTION

A primary goal in postsecondary education is to maximize students' learning outcomes and academic success, and self-assessment skills have been long associated with achieving these outcomes (Schunk, 2014; Walser, 2009). Self-assessment has been used to describe various activities (Andrade, 2019), and although recent literature has questioned the meaning and usefulness of differing self-assessment mechanisms, it is still recognized as a crucial skill for effective learning to take place, facilitate critical thinking skills, and developing lifelong learners (Andrade, 2019; Boud et al., 1989; Boud et al., 2015; Lew et al., 2015). Epstein et al. (2008) described self-assessment as on "ongoing moment-to-moment self-monitoring", where self-monitoring "refers to the ability to notice our own actions and curiosity to examine the effects of those actions" (p. 5) Additionally, a characteristic of successful learners is the ability to accurately self-assess strength and weakness (Boud et al., 1989; Boud et al., 2015). However, it is well documented that students lack the ability to accurately self-assess their perceived performance on both objective and ability-based skills, which may be contributed to lack of metacognitive awareness (Boud et al., 2015; Brown et al., 2015; Husamah, 2015; Lindsey et al., 2015; Pantiwati et al., 2017). Recent literature is conclusive in that the term *consistency* is more appropriate than *accuracy*, especially when in the classroom environment, as the quality of students' self-assessment skills is being compared to assessments by an expected standard such as their teachers (Andrade, 2019).

Over the last few decades, a variety of learning methods have been implemented in postsecondary education for purposes of developing self-assessment skills. Simulation is one example that has demonstrated effectiveness in facilitating development of self-assessment skills on ability-based performance such as safety, confidence, communication, clinical decision-making, and self-efficacy (Chamberlain, 2017; Macauley et al., 2017; Seif et al., 2012; Silberman et al., 2016). Additionally, implementing simulation into the academic curriculum is beneficial for students in facilitating metacognitive awareness (Pantiwati et al., 2017). Simulation can lead to better learning outcomes as they provide guided experiential learning that are specifically structured to facilitate essential competencies for clinical practice (Sabus et al., 2016). Simulation is driven by constructs of Social Cognitive Theory such as modeling, motivation, self-efficacy, and learner-centered

instruction. These experiences introduce active learning opportunities in conjunction with observational learning through modeled-behavior and developing higher self-efficacy and are developing self-assessment skills as well as facilitating intrinsic motivation (Brown et al., 2015; Lindsey et al., 2015).

Some post-professional healthcare programs have adopted simulation as a standard component of the program's curriculum to develop self-assessment skills, especially with ability-based performances such as professional behavior skills. The purpose of this study was to investigate if physical therapy students' self-assessment on professional behaviors skills were consistent with faculty assessment following a simulation experience. Following participation in a structured simulation experience, students filled out a standardized self-assessment instrument on professional behaviors and rated their perceived abilities on 10-professional behavior skills that have been deemed critical in the professional development process (May et al., 1995; May et al., 2010). The simulation experience was not intended to improve students' level of professionalism but instead to give students a more accurate perception of their baseline level of function on a continuum, and to facilitate metacognitive awareness.

2. METHODS

Following ethical approval from the Institutional Review Board (IRB) at Capella University and Touro College, a sample of 36 self-assessments records from a cohort of physical therapist students matriculated in n accredited Doctor of Physical Therapy (DPT) were retrieved for secondary analysis from the academic institution. The self-assessments were included if it was, (a) completed by a student in their third academic semester in the DPT program, (b) completed by a student who did not have previous clinical internships in the DPT program, and (c) completed by a student who participated in the simulation experience. Self-assessments were excluded if, (a) it was not linked to any student, (b) it contained incomplete, missing, or any sections skipped over that made the rating score not evident, (c) it was completed by a student who was dismissed from the DPT program at any point before or after the self-assessment tool was initially filled out, (d) it was completed by a student that had a previous meeting with the Professional Behaviors Committee in the DPT program, and (e) it was completed by a student that was an advisee of the faculty who served as the rater. Of the 36 records retrieved for review, 11 did not meet inclusion criteria and were excluded from statistical analysis.

2.1 Instrument

The Professional Behaviors Assessment Tool (PBAT) is a standardized self-assessment instrument that is commonly used by academic institutions across the United States for students to self-assess professional behavior. It is utilized to facilitate professional development in the clinic and classroom (May et al., 1995; May et al., 2010), and is often implemented for students when preparing for clinical internships. This self-assessment lists 10-professional behaviors, listed in order of importance: critical thinking, communication, problem solving, interpersonal skills, responsibility, professionalism, use of constructive feedback, effective use of time and resources, stress management, and commitment to learning. These behaviors have been deemed critical in accelerating the professional development process (May et al., 1995; May et al., 2010). The PBAT contains a Likert type rating scale for each of the 10-professional behaviors: *Beginner (1)*, *Intermediate (2)*, *Entry-Level (3)*, and *Post Entry-Level (4)*. Each section on the PBAT tool begins by naming and defining the professional behavior and providing specific examples of task-performance of a professional behavior at each of the four levels. This is then followed by a prompt

to reflects on each professional behavior and assigns a rating based on their perceived ability to demonstrate that professional behavior. The PBAT tool was chosen by the academic institution under study to meet curricular needs and thus was used as the outcome variable to analyze consistency between student and faculty rating.

2.1.1 Reliability and Validity

The PBAT is a reliable and valid tool to assess professional behaviors of DPT students. The reliability of the PBAT and other similar ability-based assessments have shown acceptable, good, and high inter-rater reliability (Denton et al., 2017; May et al., 2010). Denton et al. (2017) reported internal consistency using a Cronbach α range of .71 to .93. They also reported an intraclass correlation coefficient (ICC) of .75, and a test-retest reliability of .79. Yuen et al. (2016) utilized a factor analysis for selected categories of professional behaviors and reported internal reliability levels of .63 to .79. Ability-based assessments tools such as the PBAT have been found to contain strong face validity and fair construct validity (May et al., 2010; Yuen et al., 2016; Lin et al., 1991; Jette et al., 2003; Panadero et al., 2015). Jette et al. (2003) reported on the construct validity for the PBSA tool, they found scores > 0.40 using Cronbach α, which translates into "fairly well" (p. 435).

2.2 Procedure

To facilitate self-assessment skills on professional behaviors, the simulation experience contained case studies that aimed to challenge professional behaviors and was structured to mimic the clinical environment. Following participation in the simulation experience students completed the self-assessment on professional behaviors. For purposes of this research study, a full-time core faculty member of the DPT program was selected to rate each student using the same self-assessment tool and report on each student's ability in demonstrating each professional behavior listed on the PBAT. Of the 10-core faculty in the DPT program, a process of elimination was used to select one faculty who served as the rater. Factors for consideration in the process were, (a) familiarity with students in the cohort; faculty were excluded if they were not the primary professor to any classes in the first three academic semesters of the DPT program, and (b) teaching content; which classes were taught by the faculty. For example, lab intensive courses allot greater interaction between students and faculty and involve student demonstration on ability-based performance. Of the remaining faculty, years of experience teaching, role in the DPT program, and familiarity with the self-assessment instrument were considered. The faculty member who was eventually selected as the rater observed students during the simulation experience and was then asked to complete the PBAT on each student in the cohort and report the faculty's perceived rating for each professional behavior on the PBAT based on their knowledge and familiarity of the student in conjunction with their ability to demonstrate professional behaviors during the simulation experience. The faculty rater was blinded to ratings reported by students on the self-assessment. The ratings for all 10-professional behaviors were summed for each self-assessment to produce one total numerical rating, which was then paired to the corresponding numerical rating from the faculty rater.

2.3 Analysis of Data

Data were entered into SPSS statistics v26 (Amronk NY) for statistical analysis. Assumptions for Normality were statistically analyzed using the Shapiro-Wilk test and Descriptive statistics. Based on the results of normative data, a two-way paired sample *t*-test was used to

determine the difference of Means and Pearson Correlation Coefficient (r) for correlational analysis. The Mean (M), Standard Deviation (SD), Effect Size (ES), and Standard Error of Measurement (SEM) with 95% Confidence Intervals (CI) were analyzed and reported for the summed scores on the PBAT. In addition, individual analysis for each of the 10 professional behaviors on the PBAT was also completed.

3. RESULTS

The sample of 25 self-assessments that met inclusion criteria contained self-assessments from 10 students that identified as female, and 15 students that identified as male. The DPT program under study is in a suburban area in the northeast region of the United States. The DPT program had a total of 10 core faculty member, of which one was selected to assess each student in the cohort on professional behavior skills. The faculty selected to assess students on professional behaviors had over 25 years of experience as core-faculty in the DPT program, was the primary professor for two consecutive semesters of the first three academic semesters in the DPT program. Both classes taught by this faculty were lab intensive classes with practical performance as a component of the class, where demonstration of professional behaviors were also required. Lastly, the faculty selected served as the chairperson of the Professional Behaviors Committee in the DPT program, thus was familiar with the utilization of the PBAT.

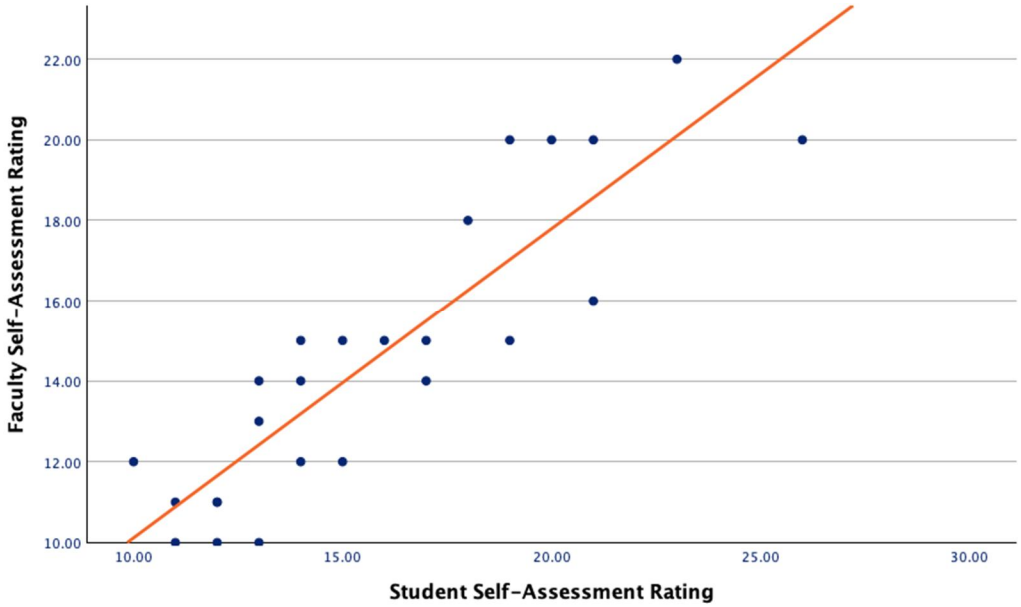
Results for Normality testing using the Shapiro-Wilk test yielded a significance level of 0.161. Table 1 demonstrates the consistencies for the self-assessment between students and faculty, which is the sum of the reported ratings for all 10-professional behaviors on the self-assessment. A paired sample *t*-test showed the M ± SD (SEM) for student reported rating was 15.84 ± 4.15 (0.83) with a 95% CI = (4.13-17.55), and faculty reported rating was 14.60 ± 3.61 (0.72) and 95% CI = (13.11-16.09), with two-tailed significance of $p = .004$ The correlation between student and faculty reported rating on the self-assessment using Pearson's Correlation Coefficient (r) = 0.884 with a two-tailed significance level $p < .001$. This analysis yielded a moderate ES Cohen's $d = .638$. Figure 1 demonstrates the correlation for student and faculty reported ratings on the self-assessment. Figure 2 exhibits a linear trend analysis demonstrating consistencies between student and faculty reported ratings on the self-assessment. Table 2 shows the consistencies of student and faculty reported ratings for each professional behavior on the self-assessment.

Table 1. Reported Ratings and Consistencies on the Professional Behaviors Assessment Tool (PBAT).

		Rater	M	SD	SEM	95%(CI)	Significance (p < .005)	Correlation Pearson (r)	Effect Size Cohen's d
Pair 1	PBAT Total Score	Student Faculty Paired Difference	15.84 14.60 1.24	4.15 3.61 1.94	.83 .72 .39	(14.13 – 17.55) (13.11 – 16.09) (0.44 – 2.04)	p = .004	.884 (p < .001)	.638

Note. This analysis consists of the summed ratings for all 10-professional behaviors.

Figure 1. Correlation of student and faculty reported ratings on the self-assessment ratings.



Note. Pearson correlation for paired ratings (r = .884, p < .001) with R^2 = .78

Figure 2. Linear trend on consistencies between student and faculty reported ratings on the self-assessment.

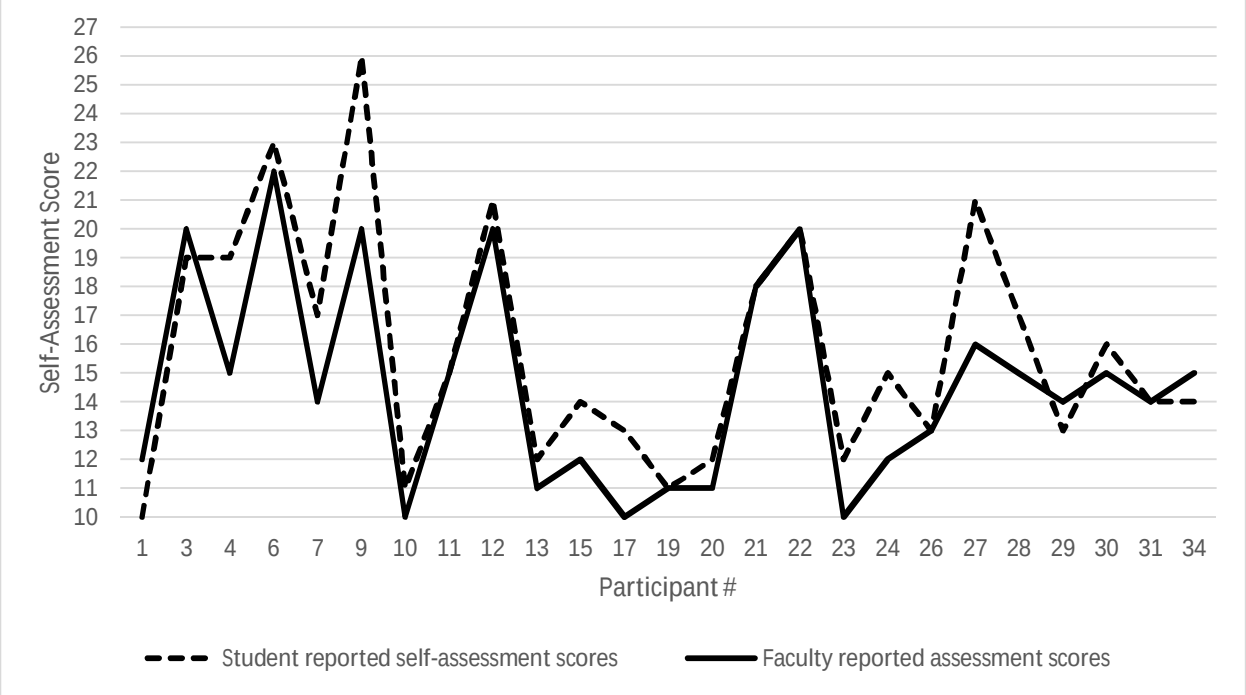


Table 2. Consistencies of student and faculty reported ratings for each professional behavior on the self-assessment.

	Behavior	Rater	M	SD	SEM	95%(CI)	Significance (p < .005)	Correlation Pearson (r)
Pair 1	Critical Thinking	Student Faculty Paired Difference	1.48 1.48 0.00	.51 .51 .50	.11 .11 .10	(-.21 - .21)	p = 1.000	.519
Pair 2	Communication	Student Faculty Paired Difference	1.64 1.52 .12	.76 .59 .53	.15 .12 .12	(-.10 - .34)	p = .265	.721
Pair 3	Problem Solving	Student Faculty Paired Difference	1.64 1.40 .24	.57 .50 .72	.12 .10 .14	(-.06 - .54)	p = .110	.088
Pair 4	Interpersonal Skills	Student Faculty Paired Difference	1.96 1.52 .44	.68 .51 .65	.14 .10 .13	(.17 - .71)	p = .002	.426
Pair 5	Responsibility	Student Faculty Paired Difference	1.44 1.56 -.12	.51 .51 .53	.10 .10 .11	(-.34 - .10)	p = .265	.461
Pair 6	Professionalism	Student Faculty Paired Difference	1.52 1.68 -.16	.51 .56 .55	.10 .11 .11	(-.39 - .07)	p = .161	.464
Pair 7	Use of Constructive Feedback	Student Faculty Paired Difference	1.80 1.40 .40	.58 .50 .58	.12 .10 .12	(.16 - .64)	p = .002	.433
Pair 8	Effective Use of Time and Resources	Student Faculty Paired Difference	1.16 1.24 -.08	.37 .44 .49	.07 .09 .10	(-.28 - .12)	p = .425	.266
Pair 9	Stress Management	Student Faculty Paired Difference	1.80 1.24 .56	.58 .44 .51	.12 .09 .10	(.35 - .77)	p < .001	.530
Pair 10	Commitment to Learning	Student Faculty Paired Difference	1.48 1.46 -.08	.59 .51 .49	.12 .10 .10	(-.28 - .12)	p = .425	.601

4. DISCUSSION AND RECOMMENDATIONS

The purpose of this study was to investigate if physical therapy students' self-assessment on professional behaviors skills were consistent with faculty assessment following a simulation experience. Professional behavior skills are a crucial component for clinical readiness and professional development (Anderson et al., 2013; May et al., 1995). Research has been conclusive that students lack self-assessment skills and do not have an accurate perception of their abilities, and there are gaps in knowledge as to why (Boud et al., 1989; Boud et al., 2015; Brown et al., 2015; Husamah, 2015; Lew et al., 2010; Lindsey et al., 2015; Panadero et al., 2015; Walser, 2009). Learning in a simulated environment has been found to assist students in developing better self-assessment skills and developing metacognition (Macauley et al., 2017; Sabus et al., 2016; Seif et al., 2012). In the current study, self-assessments were completed by students after participation in a simulation experience that aimed on professional behavior skills, and consistency was gaged through comparison to reported ratings from a faculty rater.

Data were analyzed for Normal distribution to ensure selection of the most appropriate tests to analyze the data. The analysis met assumptions for Normality, and thus a parametric test was selected to analyze the data for significance. This study found a good to excellent positive correlation between group variables, where 0-.25 = little or no correlation; .25-.50 = moderate to good correlation; .50-.75 = moderate to good correlation; and >.75 = good to excellent correlation (Portney et al., 2015). Results from this study showed that there was a strong positive correlation between students' perceptions on their abilities and their abilities perceived by faculty on professional behavior skills following a simulation experience, and Figure 2 demonstrates the consistencies between student and faculty ratings.

The central tendency and quartile hinges showed that overall, students' reported ratings were higher than faculty reported ratings, which is consistent with previous literature (Boud et al., 2013; Tejeiro, 2012). Tejeiro et al. (2012) reported that students perceive themselves as having greater abilities when compared to their professors. Also consistent with previous findings, the largest inconsistencies in reported ratings between student and faculty were at the upper quartiles indicating that students who perceived themselves as having highest baseline abilities had the greatest inconsistencies in ratings when compared to faculty. Previous literature reported consistent findings that lower performing students tend to rate themselves higher while higher perming students tend to rate themselves lower (Boud et al., 1989; Boud et al., 2013). Therefore, there may be greater inconsistencies between student and faculty ratings when the assessment was completed by a lower performing student. A limitation of this study was that detailed demographic data were reported. Collecting additional demographic data such as age, academic performance, cultural background, and previous volunteer or work experience in the clinical setting would be beneficial for future studies to provide additional insight to correlations between these variables.

Table 2 reported on consistencies for each professional behavior listed on the self-assessment. Individual analysis of professional behaviors identified with highest correlation in reported ratings between student and faculty were, communication (r = .721, p =.265), commitment to learning (r = .601, p=.425), stress management (r=.530, p= <.001), and critical thinking (r=.519, p=1.00). While the professional behaviors identified as having the lowest correlation were problem solving (r=.088, p=.110) and effective use of time and resources (r=.266, p=.425). Future studies may investigate if structuring the simulation experience with greater focus on one specific behavior is effective in improving self-assessment skills on that behavior when compared to a standard or via a pre-to-posttest analysis.

Similar studies reported correlations between student reported ratings on self-assessment and other measures to have weak to strong positive correlations ranging from r = .20 - .80 (Brown et al., 2013). However, these studies looked at summative self-assessment that make the reported correlations harder to interpret, while studies on formative assessments report increased consistency (Andrade, 2019). The current study contained a formative self-assessment.

For this study a faculty was selected to compare consistency. Some literature has questioned the reliability of teacher assessments, reporting that teachers' assessments may be inaccurate and thus should not be used as an external judgement to determine consistency (Admiraal et al., 2015; Baxter et al., 2011; Brown et al., 2015; De Grez et al., 2012; Falchikov, 2005; Leach, 2012). However, other literature argues that this assumption is multifactorial and can be consistent under certain circumstances. First, when comparing to external judgement greater inconsistencies are reported for summative assessments, while formative self-assessments that have specified guidelines or a rubric are more consistent (Admiraal et al., 2015; Baars et al., 2014; Baxter et al., 2011; Bol, et al., 2012; De Grez et al., 2012; Panadero et al., 2014;). Second, greater consistencies between student and teacher ratings are noted if the self-assessment is not a graded component (Tejeiro et al., 2012). Third, greater consistencies are reported or if the learner older and academically competent (Hacker et al., 2000; Lew et al., 2010; Alaoutinen, 2012; Guillory and Blankson, 2017; Butler, 2018; Nagel and Lindsey, 2018). Thus, greater consistencies are reported for students in higher education (Ratminingsih, 2018) while studies on children report less consistencies (Harris et al., 2013; Bourke, 2016). A last argument that should not be ignored is the degree to which the teacher knows the students. In DPT programs, a cohort of students spend significantly more time in the classroom with any one given instructor, and it is typical that the same instructor teaches the same cohort of students in multiple classes and classes that have a laboratory or practical component where students get to demonstrate skills to the instructor in the psychomotor, cognitive, and affective domains.

The current research study found a large positive correlation that was statistically significant ($r = .884$, $p < .001$), indicating consistencies between student and faculty reported ratings that had a mean difference that was statistically significant. Individual analysis of each professional behavior showed that 7 of the 10 professional behaviors had a significance level $p > .005$. This indicates the Mean difference of the reported ratings for those professional behaviors were not significantly different; there were no statistically significant difference in reported ratings between student and faculty.

It is recommended that future studies implement a pretest-posttest research design, if feasible, to gain additional insight to the magnitude of calibration that may be occurring due to patient simulation. In this scenario, participants would complete the self-assessment instrument prior to and after patient simulation, and analysis would be performed on pretest ratings paired with posttest ratings, rather than a comparison to faculty ratings. However, simulation is now well documented in its effectiveness with facilitating development of professional behaviors and other important clinical skills such as safety, confidence, communication, clinical decision-making, and self-efficacy (Chamberlain, 2017; Seif et al., 2016; Silberman et al., 2016). It is also reported that implementing patient simulation experiences into the didactic curriculum is beneficial for students in facilitating metacognitive awareness to their abilities with clinical skills such as communication, collaboration, and safety awareness (Pantiwati et al., 2017), which can result with inaccurate perceptions for self-abilities, and difficulty identifying the level of function in the continuum of professional development.

5. CONCLUSIONS

Inconsistencies in student self-assessment may be contributed to lack of metacognitive awareness (Pantiwati et al., 2017). Simulation activities can help develop student self-assessment skills on professional behaviors. This study found strong consistencies between student and faculty reported ratings on self-assessment of professional behaviors following a simulated experience that aimed at professional behavior skills, suggesting that structured simulated experiences may play a role of developing self-assessment skills on professional behaviors for physical therapy students. However, further research is needed as to why students lack consistency on self-assessment and how to develop those skills.

References

Admiraal, W., Huisman, B., and Pilli, O. (2015). Assessment in massive open online courses. *The Electronic Journal of E-Learning*, *13*(4), 207–216.

Alaoutinen, S. (2012). Evaluating the effect of learning style and student background on self-assessment accuracy. *Computer Science Education,* 22(2), 175–198. https://doi.org/10.1080/08993408.2012.692924

Anderson, D. K., & Irwin, K. E. (2013). Self-assessment of professionalism in physical therapy education. *Work*, *44*(3), 275-281. https://doi.org/10.3233/WOR-121504

Andrade, H. (2019). A critical review of research on student self-assessment. Frontiers in Education: Assessment, Testing, and Applied Measurement, 4(87), 1-13. https://doi.org/10.3389/feduc.2019.00087

Baars, M., et al. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Journal of Learning and Instruction, 33*, 92–107. https://doi.org/10.1016/j.learninstruc.2014.04.004

Baxter, P., and Norman, G. (2011). Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. *Journal of Advanced Nursing, 67*, 2406–2413. doi: 10.1111/j.1365-2648.2011.05658.xBoud et al., 1989

Bol, L. et al. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*(4), 280–287. https://doi.org/10.1016/j.cedpsych.2012.02.004

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529-549. http://www.jstor.org/stable/344739

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, *38*(8), 941-956. https://doi.org/http://dx.doi.org/10.1080/02602938.2013.769198

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher Education Research and Development*, *34*(1), 45-59. https://doi.org/10.1080/07294360.2014.934328

Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge Journal of Education, 46,* 97–111. https://doi.org/10.1080/0305764X.2015.1015963

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy and Practice*, *22*(4), 444-457. https://doi.org/10.1080/0969594X.2014.996523

Brown, G. T., and Harris, L. R. (2013). "Student self-assessment," in *Sage Handbook of Research on Classroom Assessment,* ed J. H. McMillan (Los Angeles, CA: Sage), 367–393. doi: 10.4135/9781452218649.n21

Butler, D. L., and Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Rev. Educ. Res.* 65, 245–281. doi: 10.3102/00346543065003245

Chamberlain, J. (2017). The impact of simulation prebriefing on perceptions of verall effectiveness, learning, and self-Confidence in nursing students. *Nursing Education Perspectives (Wolters Kluwer Health)*, *38*(3), 119-125. https://doi.org/10.1097/01.NEP.0000000000000135

De Grez, L., Valcke, M., and Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*, 129–142. doi: 10.1177/1469787412441284Denton et al., 2017

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA: Journal of the American Medical Association, 287*(2), 226–235. https://doi.org/10.1001/jama.287.2.226

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395-430. https://doi.org/10.2307/1170205

Guillory, J. J., & Blankson, A. N. (2017). Using recently acquired knowledge to self-assess understanding in the classroom. *Scholarship of Teaching and Learning in Psychology, 3*(2), 77–89. https://doi.org/10.1037/stl0000079

Hacker, D. J. et al. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160–170. https://doi.org/10.1037/0022-0663.92.1.160

Harris, L. R., and Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: case studies into teachers' implementation. *Teaching & Teacher Education, 36*, 101–111. https://doi.org/10.1016/j.tate.2013.07.008

Husamah, H. (2015). Blended project based learning: Metacognitive awareness of biology education. *Journal of Education and Learning*, *9*(4), 274. https://doi.org/10.11591/edulearn.v9.i4.2121

IBM Corp. Released 2019. *IBM SPSS Statistics for Macintosh*, Version 26.0. IBM Corp.

Jette, D., Portney, L. (2003). Construct validation of a models for professional behavior in physical therapist students. *Physical Therapy Journal 83*(5), 432-443.

Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment and Evaluation in Higher Education, 37*, 137–147. doi: 10.1080/02602938.2010.515013Lew et al., 2015

Lew, M., Alwis, W., & Schmidt, H. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education 35*(2), 135-156. https://psycnet.apa.org/doi/10.1080/02602930802687737

Lin, R. L., Dunbar, S., & Baker, E. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* 20(8), 15-21. https://doi.org/10.3102/0013189X020008015

Lindsey, B. A., & Nagel, M. L. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. *Physical Review Special Topics - Physics Education Research, 11,* 1-11. https://doi.org/10.1103/PhysRevSTPER.11.020103

Macauley, K., Brudvig, T., Kadakia, M., & Bonneville, M. (2017). Systematic review of assessments that evaluate clinical decision making, clinical reasoning, and critical thinking changes after simulation participation. *Journal of Physical Therapy Education, 31*(4), 64-75. https://doi.org/10.1097/JTE.0000000000000011

May, W., Kontney, L., & Iglarsh, A. (2010*). Professional behaviors for the 21st century assessment.* Marquette University. https://www.marquette.edu/physical-therapy/information-clinical-facilities.php

May, W., Morgan, B., Lemke, J. C., Karst, G., & Stone, H. (1995). Model for ability-based assessment in physical therapy education. *Journal of Physical Therapy Education, 9,* 3–6.

Nagel, M., and Lindsey, B. (2018). The use of classroom clickers to support improved self-assessment in introductory chemistry. *Journal of College Science Teaching, 47*(5), 72–79. https://www.nsta.org/store/product_detail.aspx?id=10.2505/4/jcst18_047_05_72

Panadero, E., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education, 21*(2), 133–148. https://doi.org/10.1080/0969594X.2013.877872

Panadero, E., Tapia, J., & Huertas, J. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning, and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806-813. https://doi.org/10.1016/j.lindif.2012.04.007

Pantiwati, Y., & Husamah, H. (2017). Self and peer assessments in active learning model to increase metacognitive awareness and cognitive abilities. *International Journal of Instruction, 10*(4), 185-202. https://doi.org/10.12973/iji.2017.10411a

Portney, L. G., & Watkins, M. P. (2015). *Foundations of clinical research: Applications to practice.* F.A. Davis.

Sabus, C., & Macauley, K. (2016). Simulation in physical therapy education and practice: Opportunities and evidence-based instruction to achieve meaningful learning outcomes. *Journal of Physical Therapy Education, 30*(1), 3-13. https://doi.org/10.1097/00001416-201630010-00002

Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2014). *Motivation in education: Theory, research and application* (4th ed.). Pearson Education.

Seif, G. A., Brown, D., & Annan-Coultas, D. (2013). Video-recorded simulated patient interactions: Can they help develop clinical and communication skills in today's learning environment? *Journal of Allied Health 42*(2): e37-e44.

Silberman, N., Litwin, B., Panzarella, K., & Fernandez-Fernandez, A. (2016). High fidelity human simulation improves physical therapist student self-efficacy for acute care clinical practice. *Journal of Physical Therapy Education, 30*(1), 14-24. https://doi.org/10.1097/00001416-201630010-00003

Tejeiro, R. A., Gomez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., and Emberley, E. (2012). Summative self-assessment in higher education: implications of its counting towards the final mark. *Electronic Journal of Research in Education and Psychology, 10*, 789–812.

Walser, T. M. (2009). An action research study of student self-assessment in higher education. *Innovative Higher Education, 34*(5), 299-306. https://doi-org.library.capella.edu/10.1007/s10755-009-9116-1

Yuen, H., Azuero, A., Lackey, K., Brown, N., & Shrestha, S. (2016). Construct validity test of evaluation tool for professional behaviors of entry-level occupational therapy students in the united states. *Journal of Educational Evaluation for Health Professions, 13*(22).