▲ A Method for Assessing Reflective Journal Writing

Margaret M. Plack, PT, EdD
Maryanne Driscoll, PhD
Sylvene Blissett, PT, MPH, DPT, PCS
Raymond McKenna, PT, PhD
Thomas P. Plack, PT, DPT

Reflection is widely accepted as a learning tool and is considered integral to professional practice. Journal writing is advocated in facilitating reflection, yet little is written about how to assess reflection in journals. The purpose of this study was to develop and test a method of assessing the elements of reflection in journals and to determine whether, and to what level, reflection occurs in journals. Twenty-seven physical therapy students maintained written reflective journals throughout three of their four eight-week clinical affiliations. The students were introduced to concepts of reflective practice with definitions of terms and reflective questions before their second affiliation. A coding schema was developed to assess the journals. Three raters assessed forty-three journals. The text of each journal was analyzed for evidence of nine elements of reflection, and each journal was categorized as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection. Descriptive statistics were used to demonstrate evidence of reflection. Reliability between each pair of raters was assessed using percent agreement, ϕ coefficients, and γ statistics. Interrater reliability of all raters was assessed using intraclass correlation coefficients (ICC[2,1]). Results showed that the raters assessed 95.3%–100% of the journals as showing at least one element of reflection. The percent agreement between rater pairs for the nine elements of reflection ranged from 65.1% to 93.0%, the ϕ coefficient ranged from 0.08 to 0.81, and the ICC(2,1) values used to assess reliability among the three raters on each element ranged from 0.03 to 0.72. Averaging the assessment of the three raters for the overall journal, 14.7% of the journals were assessed as showing no evidence of reflection, 43.4% as showing evidence of reflection, and 41.9% as showing evidence of critical reflection. The percent agreement between rater pairs for the overall assessment of the journals ranged from 67.4% to 85.7%, the γ statistic ranged from 0.88 to 0.98, and the ICC(2,1) among all raters was 0.74 (95% confidence interval, 0.61–0.84). These results represent an acceptable level of agreement for use of this method of assessment for educational purposes. The coding schema developed provides a mechanism to assess evidence of reflection in written journals, which will enable instructors to evaluate student competency, obtain a baseline for facilitating reflective practice, and assess their own efficacy in facilitating reflection among students. J Allied Health 2005; 34:199-208.

REFLECTION, as a method of learning from experience, has been widely accepted in various domains of professional education, particularly in teacher education and nursing. In the medical and allied health professions, it has been considered critical to professional practice, leading certain accrediting bodies to expect education programs to prepare students to be reflective practitioners.[1] Much has been written about reflective practice, particularly with respect to how it can be incorporated into the classroom. Further, journal writing has been suggested as a mechanism to facilitate reflection.[2–12] However, little has been written about how to assess reflection, particularly from the perspective of assessing its processes or components evident within journals.[10–14] Without a mechanism of assessment, educators cannot fully determine the extent to which journal writing truly facilitates reflective practice. The purpose of this study was to develop and test a method of assessing reflection in journals and to determine whether, and to what level, reflection occurs in student journals.

Literature Review

The terms "reflection" and "reflective practice" have become commonplace; when discussed, many clinicians espouse that they reflect routinely. The question raised by these researchers was "while many say they reflect, do they?" To fully investigate this question, the researchers explored the theoretical literature to define reflection and its importance in the learning process as well as current literature to determine how reflection has been measured.

Dr. M. Plack is Assistant Professor and Chair, Department of Health Care Sciences, School of Medicine and Health Sciences, George Washington University, Washington, DC; **Dr. Driscoll** is Associate Professor and **Dr. Blissett** is Academic Coordinator of Clinical Education, Physical Therapy Program, Touro College, New York, New York; **Dr. McKenna** is Associate Professor, Physical Therapy Program, Stony Brook University, Stony Brook, New York; and **Dr. T. Plack** is a clinician in Washington, DC.

Received June 7, 2004; revision accepted November 18, 2004.

Address correspondence and reprint requests to: Margaret M. Plack, PT, EdD, Program in Physical Therapy, School of Medicine and Health Sciences, George Washington University, 900 23rd Street, NW, Sixth Floor, Washington, DC 20037. Tel 202-994-7763; fax 202-994-8400; e-mail mppt@aol.com.

Reflection: What Is It?

Reflection is a generic term that has many different definitions. Boyd and Fales[15] defined reflection as the process of examining an experience that raises an issue of concern. They described it as an internal process that helps the individual refine his or her understanding of an experience, which may lead to changes in the individual's perspective. Boud et al.[16] defined reflection as the cognitive and affective behaviors individuals engage in that result in new insights and deeper understandings of their experiences. Schön[17] described how clinicians, faced with problems, are often forced to stop, think, and problem solve on the spot. This is what Schön termed "reflection in action." He further posited that reflective practitioners revisit their experiences through reflection on action, which occurs after action when practitioners attempt to view problems from alternative perspectives. Killion and Todnem[18] extended Schön's work as they discussed the importance of reflection for action or being able to anticipate future problems and being able to modify actions to improve future practice. Mezirow[9] discussed the content or focus of the reflective process, stating that reflection is not simply stopping to think and problem solve or plan for future action based on what you already know; rather, it is critically questioning the content, process, and premise underlying the experience in an attempt to make meaning or better understand the experience. Mezirow[9] contended that reflection is a higher order, conscious thought process that enables one to begin to correct assumptions and distorted beliefs that may lead to revised interpretations of one's experiences and ultimately to new behaviors (i.e., transformational learning).

While there are many definitions for reflection, Atkins and Murphy[19] maintained that there are essential elements to this process. They noted that in the reflective process, an awareness of uncomfortable feelings and thoughts (i.e., trigger event) is followed by a critical analysis of both the feelings and the experience. Atkins and Murphy[19] asserted that it is through this analytical process that the reflective practitioner develops a change in perspective. Mezirow[9] claimed that the reflective process leads to a more open, discriminative, and inclusive perspective or worldview.

Reflection: Why Is It Important?

Reflection gives meaning to experience; it turns experience into practice, links past and present experiences, and prepares the individual for future practice. It is the "hallmark of professional behavior."[20] Schön[17] stated that curricula in professional programs often favor technical rationalism (i.e., knowledge and skills) over problem solving and professional development. He believed that the reflective thinking process is critical to problem solving and to professional practice. Reflection is particularly important in physical therapy and other allied health fields where development of a professional identity goes beyond technical knowledge and skill to more abstract constructs such as critical thinking and professional values, attitudes, and beliefs. Further, reflection promotes a deep approach to learning and fosters lifelong learning as students learn to reframe problems, question their own assumptions, and attend to their own learning needs.[9,21] Journal writing provides a mechanism for individuals to describe their experiences and begin to use the reflective and analytical or critical thinking processes to extract deeper meaning from those experiences.[2,5,6,8,9,11,21-23]

Brookfield[24] defined critical thinking as a direct outcome of the reflective process described by both Mezirow[9,25] and Schön.[17,26] He described how thinking critically fosters the recognition of assumptions, beliefs, and values that underlie our thought processes as we solve problems, anticipate outcomes, and justify our actions.[24] Critical thinking is achieved by recognizing and analyzing multiple perspectives and is essential to problem solving.

Reflection: Why Measure It?

As noted, it is generally accepted that reflective practice is critical to professional development. Journals have been used in physical therapy[6,10,11,14,27] and in other disciplines to promote reflection among students.[2,4,5,12,21,28,29] However, maintenance of reflective journals does not necessarily ensure that students reflect; rather, some may simply describe their experiences and not take the critical step toward analysis as proposed by Atkins and Murphy.[7,8,19,30] In keeping with the move toward evidence based practice, Pee et al.[21] noted that assessment of the efficacy of this strategy in promoting reflection is essential.

Assessment of the products of reflective journals, however, is controversial.[2,3,7,31] Some educators suggest that reflective journals not be evaluated because they contend that evaluation may impact the content of the reflection.[2,7,24,31,32] Others question how to effectively and consistently evaluate what many consider personal thoughts and feelings.[30,33,34] Still others question the validity of assessing subjective knowledge.[3] While noting these obstacles and acknowledging that placing judgment on what students write in journals could potentially impact their writing, if one is to develop reflective practitioners, the need for a mechanism to assess a student's capacity to reflect remains. To mitigate some of these obstacles, Bourner[3] proposed separating content and process in the assessment of journal writing. This would allow for assessment of the reflective process without placing judgment on the subject nature of the reflection.

Assessment of reflective journals is important for a number of reasons. In developing reflective practitioners, it is essential to determine whether students have developed proficiency in the reflective process. Moreover, assessment allows instructors to provide feedback to students on their learning. Further, as Bourner notes, "most students pay most attention to what is assessed. That which is not assessed is

most likely to be most neglected."[3] The assessment process also provides feedback to faculty members both about the efficacy of their teaching strategies and about achievement of course and curricular outcomes related to the reflective process.[3] Finally, assessment of the products of reflection can be used to demonstrate achievement of curricular outcomes for the purpose of accreditation.

Reflection: How Do We Measure It?

While many advocate journal writing to promote reflective practice, little has been written about how to determine whether students show evidence of reflection in their writing.[6,10–14] A number of qualitative studies have examined the content of reflective writing.[6,11,12,22,35] Jensen and Denton[6] and Williams et al.[11] used qualitative methods to analyze journals from physical therapy students, Kalliath and Coghlan[22] studied students of organizational development, and Drevdahl and Dorcy[35] studied nurses. Each reported themes that emerged from the content of these journals, which broadly included communication, interpersonal interactions, personal growth, and professional knowledge and skills.[6,11,22,35] While themes regarding the content of the reflective products were presented, what was not evident in these studies was what Pee et al.[21] questioned, which is whether students reflected on a superficial, more descriptive level or a deeper, more analytical level. Thus, no judgment could be made regarding the students' proficiency in the process of reflection.

Other researchers used quantitative methods to analyze the processes of reflection present within reflective journals. In a study with student nurses, Wong et al.[12] used the reflective processes as defined by Boud et al.[16] to develop a mechanism to analyze journals in terms of the levels of reflection present. In addition, they categorized each as nonreflector, reflector, or critical reflector, based on the work of Mezirow.[9] Wong et al.[12] reported reliable results (r = 0.88) using Mezirow's[9] theoretical framework, but not the framework of Boud et al.[16] Williams et al.[10] adapted the criteria from the work of Boud et al.,[16] creating a six-point ordinal scale to assess the level of reflection present in physical therapy student journals. These researchers reported an interrater reliability of 0.68 (95% confidence interval, 0.49–0.87). Finally, Pee et al.[21] used structured worksheets to assess whether dental therapy students showed evidence of reflection in their writing. These researchers reported that 86% (n = 12) of the students showed evidence of descriptive or superficial reflection, while 64% (n = 9) showed evidence of a deeper level of reflection. Interrater agreement reported in this study was 72% when rating whether the structured question had been answered and 86% when determining the level of reflection.

In summary, while reflection is generally accepted as a critical component of learning from experience and essential to professional education, limited research has been conducted to address the issue of how to assess reflection. Journals have been used in many areas of professional development, including higher education,[5,22,30] nursing,[12] allied health,[6,10,11,14,29] medicine and dentistry,[21] and adult education.[7-9,24,31] While some suggest that evaluating journals may impact what an individual chooses to write, the current researchers submit that without some mechanism of assessment, an instructor cannot determine whether students are truly competent in the skills necessary for reflective practice.

Bourner[3] suggested separating content from process in assessment to minimize the impact on the content of the journal writing. Boud et al.[16] provided a framework for understanding the stages of reflection, Mezirow[9] focused on components of the reflective process; and Schön[17] provided a time dimension. To assess competence in the reflective process, the current researchers contend that all three aspects of reflection must be evaluated to fully determine the degree to which students have mastered the elements of reflection and the reflective process. Therefore, the purpose of this study was to utilize the work of these three major theorists to develop a method to reliably assess whether students show evidence of the various elements of reflection in journals and, if present, to what level did they evidence reflection.

Methods

Instrument

Based on the theoretical and scientific literature, an instrument was designed to assess individual elements of reflection and to judge the overall level of reflection evident within each journal (Table 1). Codes and operational definitions were developed from the theories of Boud et al.,[16] Mezirow,[9] and Schön.[17] Two levels of coding were defined. Coding first occurred at the level of the words, sentences, and paragraphs within the journals. The following nine elements of reflection were operationally defined: reflection in action, reflection on action, reflection for action, content reflection, process reflection, premise reflection, returns to experience, attends to feelings, and evaluation of the experience.[9,16,17] In addition to the nine elements of reflection, each journal was coded in total and categorized as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection based on definitions from the work of Mezirow[9] and others.[12,17,24,36] Operational definitions were refined and agreed on by all three investigators before analyzing the journals (Table 1 and Appendix).

Subjects

Twenty-seven physical therapy students from the class of 2002 from Touro College were asked to maintain written journals during three of their eight-week clinical education experiences as a requirement of each clinical education unit, beginning with their second affiliation.

TABLE 1. Rating Checklist for Reflective Journals

	Code		Brief Definition*	Presence (Provide Evidence)	Comments
Level I		Element of reflection			
Time	R-I-A	Reflection in action	Occurs while in the midst of an action; on-the-spot decisions		
	R-O-A	Reflection on action	Occurs after the action has been completed		
	R-F-A	Reflection for action	Occurs before being faced with the situation; begins to plan for the future		
Content	CON	Content	Explores the experience from a number of perspectives (beyond description)		
	PROC	Process	Describes the strategies used or available		
	PREM	Premise	Recognizes and explores own assumptions, values, beliefs, and biases		
Stage	RETRN	Returns to experience	Describes the experience		
	ATTEND	Attends to feelings	Acknowledges and begins to work with feelings		
	RE-EVAL	Reevaluates	Reappraises the situation *vis-à-vis* past experiences		
Level II		Level of reflection			
Journal	NR	Nonreflection	No evidence of reflection is present		
	R	Reflection	Evidence of reflection is present		
	CR	Critical reflection	Evidence of critical reflection is present		

Note. Select all that apply.
*For complete operational definitions, see Appendix.

RATERS

Three raters, from an initial pool of four, participated in this study. Raters included (1) the director of a transitional doctor of physical therapy program with a doctor of education degree in adult education, (2) a faculty member in a professional physical therapy program with a background in education and psychology, (3) a clinician in a transitional doctor of physical therapy program, and (4) an academic coordinator of clinical education (ACCE) for a professional physical therapy program. While all four raters were expected to participate, due to time constraints, one of the four raters (the ACCE) was not available for the initial training sessions, during which the raters worked to fully understand and refine the operational definitions and rating checklist that was developed. While the first rater had formally studied principles of reflection and incorporated them into the classroom setting, all three raters were novices in assessing journals.

PROCEDURES

Before the students' second affiliation in preparation for clinical practice, the investigators spent a 1.5-hour session discussing reflective practice. The students received instruction on the importance of reflective practice in physical therapy and the need to make consistent journal entries throughout their clinical experiences to aid in the process of reflection. In addition, they were given definitions of terms and questions to be used as "reflection starters" only if they were having difficulty initiating their journaling process. The journals, together with other written assignments, were submitted to the ACCE at the end of each clinical education experience during the 2001–2002 academic year. Journals were submitted as a requirement of the experience; however, they were neither graded nor evaluated. Journals were maintained on file in the office of the ACCE. Students who submitted these journals graduated in 2002, and institutional review board approval was obtained before analyzing data.

A retrospective study was designed using these journals to develop a method of assessing the presence or absence of the nine elements and three levels of reflection present within the journals. Using the codes and operational definitions developed, raters performed a content analysis of each journal.[37] Data were subjected to two levels of coding. Coding first occurred at the level of the text (i.e., words, sentences, and paragraphs) to determine the elements of reflection present. The raters searched for the presence of the nine elements of reflection. Once evidence of the element was found, the raters were not required to continue searching for additional evidence of that particular element. In addition, it was understood by the raters that a single written passage might be rated as showing evidence of more than one element of the reflective process. Once analysis for the elements was completed, the journal was then categorized as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection as operationally defined.

In preparation for rating the journals and as a means of refining the codes, the three raters used the codes to independently rate reflective passages unrelated to the student journals. Disagreement led to discussion and further refinement of the codes, and consensus was reached on all texts. Before analysis, institutional review board approval was obtained and journals were deidentified. Once approved, a pilot study using the same three raters to rate five randomly selected journals commenced. After rating these five journals, raters again met to discuss differences. Having achieved consensus on operational definitions, the raters independently assessed the additional 43 journals.

Data Analysis

Descriptive statistics were used to quantify the elements and level of reflection present within each journal. Interrater reliability was assessed for each individual element of reflection and for the overall level of reflection evident within the journals. Inferential statistics were used to determine interrater reliability. An overall score was not assessed because the researchers were most concerned with assessing the individual elements of reflection; moreover, the elements are distinctly different and not necessarily correlated. Reliability between each pair of raters, for each element of reflection, was assessed with a measure of agreement (i.e., percent agreement) and a measure of association (i.e., ϕ coefficient). Percent agreement is the simplest and most intuitive form of interrater reliability; however, it does not consider chance agreement. The ϕ coefficient is appropriate for measuring the association between two dichotomous variables; it is a special case of the product-moment correlation coefficient.[38] Interrater reliability among all three raters for each element of reflection was assessed using intraclass correlation coefficients (ICC[2,1]).[37–39] ICCs account for both relative and absolute reliability or agreement and for association and enable the researchers to assess reliability among more than two raters for each element without distortion to nominal scale data.[37,40,41] Model 2 was used because each element was assessed by all raters who are representative of a larger population of raters (i.e., instructors), and the expectation is that the results may be generalized to other raters with similar characteristics.

Reliability between rater pairs for the overall assessment of each journal as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection was assessed using percent agreement and γ statistics. The γ statistic is appropriate for measuring the association between two variables on an ordinal scale.[42] The assessment of no evidence of reflection, evidence of reflection, or evidence of critical reflection was judged as being on an ordinal scale by the researchers. Interrater reliability among all three raters was assessed using ICC(2,1).

In establishing criteria to judge the reliability of this method of assessment, it was determined that there is no hard and fast rule regarding how high a reliability coefficient for a measure must be for it to be considered reliable. Many measurement texts do not establish acceptable reliability values, but rather suggest guidelines and note that it is dependent on how the measure is used.[37,38,43] Several issues must be considered in determining acceptable levels of reliability, such as the types of decisions to be made in using the assessment method, the potential consequences of these decisions, the number of cases used to assess the method, the specific comparisons to be made, and whether the decisions are made at an early stage, are reversible, and/or are confirmable by other data.[37,43-45] In the case of the reflective journals, the purpose of the assessment is to provide an early baseline for the facilitation of the reflective process and the results have minimal consequences, which can be confirmed through other means such as additional writings, probing questions in class, and so on. In addition, the writing in each journal is individualized, making the subject of the writing quite varied, which requires a degree of interpretation to assess. Thus, while each code/element is operationally defined, the concepts being assessed are abstract, making perfect agreement more difficult and perhaps unrealistic.[12,37] The researchers considered the type and purpose of the method of assessment and decided to use the criteria established by Rea and Parker[46] to judge reliability. These researchers indicated that ϕ coefficient values of –0.20 to ≤0.40 represent a moderate association, 0.40 to ≤0.60 represent a relatively strong association, 0.60 to ≤0.80 represent a strong association, and 0.80 to ≤1.00 represent a very strong association.[46]

Results

Twenty-seven students submitted a total of 48 journals. Five randomly selected journals from four individuals (three women and one man) were used in the initial piloting and refinement of the instrument. Results from the analysis of the remaining 43 journals are reported here. The students included 13 women and 14 men with a mean age

TABLE 2. Number of Journals Demonstrating Evidence of Each Reflective Component as per Each Rater

Reflection Component	Rater 1	Rater 2	Rater 3	Mean ± SD
Reflection in action	17	28	25	23.3 ± 5.7
Reflection on action	35	37	41	37.7 ± 3.1
Reflection for action	25	31	27	27.7 ± 3.1
Content	34	32	39	35.0 ± 3.6
Process	38	38	37	37.7 ± 0.6
Premise	17	19	19	18.3 ± 1.2
Returns to experience	39	36	43	39.3 ± 3.5
Attends to feelings	37	34	43	38.0 ± 4.6
Reevaluates	35	23	38	32.0 ± 7.9

Note. Forty-three journals were rated.

of 27.1 ± 3.6 yrs and a range of 22.3–41 yrs. Twenty-two of the journals were obtained from students on their second clinical education experience, 24 from students on their third clinical education experience, and two from students on their fourth clinical education experience.

At least one element of reflection was identified by all raters in 95.3% of the 43 journals analyzed. The number of journals that showed evidence of each element of reflection as determined by each rater appears in Table 2. Averaging these values over the three raters, the number of journals that showed evidence of each individual component of reflection assessed ranged from 18.3 ± 1.2 to 39.3 ± 3.5, with "premise reflection" noted least often and "return to experience" noted most often.

Percent agreement and ϕ coefficients used to assess interrater reliability between each pair of raters are presented in Table 3. The percent agreement between rater pairs for the nine elements of reflection ranged from 65.1% to 93.0%, and the ϕ coefficient ranged from 0.08 ± 0.18 to 0.81 ± 0.09 (Table 3). ϕ coefficients for four rater pairs could not be calculated (identified in Table 3 as "NA") due to an absence of variability in a single rater's assessment of two reflective components. Rater 3 found evidence of "return to experience" and "attends to feelings" in all 43 journals. Similarly, the lowest ϕ coefficients can be explained by the small number of agreements for the absence of a component. For example, the components with the two lowest ϕ coefficients (i.e., rater 1 vs. rater 2 for "process" and "returns to experience") had only a single agreement for the absence of the component.

The ICC(2,1) values used to assess interrater reliability among the three raters ranged from 0.03 to 0.72 and can be found in Table 4. Like the ϕ coefficient, intraclass correlation coefficients are sensitive to a lack of variability in the data. If the components "returns to experience" and "attends to feelings" are removed from the analysis, the lower bound of the range changes, resulting in a range of 0.41–0.72.

Raters made an overall assessment of each journal as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection. Taking the mean of the three raters, 14.7% of the journals were assessed as showing no evidence of reflection, 43.4% as showing evidence of reflection, and 41.9% as showing evidence of critical reflection. The percent agreement between rater pairs for the overall assessment of the journals ranged from 67.4% to 85.7%, the γ statistic ranged from 0.88 ± 0.07 to 0.98 ± 0.02, and the ICC(2,1) among all raters was 0.74 (95% confidence interval, 0.61–0.84).

Discussion

In assessing the interrater reliability of the categorization of the overall journals as showing no evidence of reflection, evidence of reflection, or evidence of critical reflection, it was determined that percent agreement between rater pairs for the overall assessment of the journals (which ranged from 67.4% to 85.7%) along with γ statistics (which ranged from 0.88 to 0.98) and the ICC(2,1) of 0.74 represent strong association and a high level of reliability. This was consistent with the work of Wong et al.,[12] who reported a reliability of 0.88.

In determining the reliability of the assessment of the presence of the nine elements of reflection, statistical values varied. Percent agreement on all nine elements and between all three pairs of raters ranged from 65.1% to 93.0%. In some instances, while percent agreements were high, the corresponding ICC values and ϕ coefficients were low. For example, while the percent agreements for "returns to experience" and "attends to feelings" were high (range, 79.1-90.7%), the ICC for each was quite low (0.03 and 0.28, respectively) and ϕ coefficients could not be determined. In reviewing the raw data and analyzing the statistical calculations, it was evident that the statistical values were particularly sensitive to the lack of variability in the data and the distribution of agreements and/or disagreements. Moreover, analysis of the raw data showed that rater 3 rated all 43 journals as having demonstrated "returns to experience," while the other two raters only indicated evidence of this element if the writer described the experience in some detail. Postanalysis discussion resulted in consensus that the interpretation of rater 1 was most theoretically correct. Similarly, the element "attends to feelings" showed very little variability, with rater 3 again rating all 43 journals as showing evidence of this element, while raters 1 and 2 rated significantly fewer journals as showing evidence of this element (i.e., 37 and 34, respectively). Postanalysis discussion revealed that rater 3 credited any mention of emotion in the journals as showing evidence of the presence of this element, while raters 1 and 2 required not only mention of emotion but also some discussion of the impact of the situation on emotion or vice versa. Consensus was that the interpretation of raters 1 and 2 was more theoretically accurate.

Using the criteria established to judge reliability, acceptable levels of interrater reliability were achieved in all cases except "returns to experience" and "attends to feelings." The

TABLE 3. Percent Agreement and Φ Coefficients and Asymptotic Standard Error for Each Pair of Raters

	Rater					
	1 vs. 2		1 vs. 3		2 vs. 3	
Reflection Component	% Agree	r^2	% Agree	r^2	% Agree	r^2
Reflection in action	69.8	0.49 ± 0.11	81.4	0.69 ± 0.08	79.1	0.57 ± 0.13
Reflection on action	86.0	0.60 ± 0.16	86.0	0.46 ± 0.15	86.0	0.20 ± 0.21
Reflection for action	76.7	0.52 ± 0.13	86.0	0.71 ± 0.11	81.4	0.59 ± 0.13
Content	90.7	0.75 ± 0.12	88.4	0.62 ± 0.13	83.7	0.55 ± 0.12
Process	81.4	0.10 ± 0.18	93.0	0.69 ± 0.17	88.4	0.48 ± 0.20
Premise	86.0	0.72 ± 0.11	90.7	0.81 ± 0.09	81.4	0.62 ± 0.12
Returns to experience	79.1	0.08 ± 0.18	90.7	NA	83.7	NA
Attends to feelings	88.4	0.62 ± 0.15	86.0	NA	79.1	NA
Reevaluates	72.1	0.51 ± 0.09	93.0	0.76 ± 0.12	65.1	0.39 ± 0.09

NA, no Φ coefficient was computed because rater 3 had a constant rating for all journals.

researchers assert that this method of assessing reflective journals is reliable for educational purposes in determining competence in the reflective process and in determining a baseline from which to begin facilitating the reflective process in students. Postanalysis discussion suggested that further refinement of the operational definitions, particularly with respect to "returns to experience" and "attends to feelings," should minimize future misinterpretations. In addition, the researchers contended that continued practice with application of the operational definitions to additional journals would enhance the reliability of this method.

Reflection can be viewed from a number of perspectives. Understanding the multiple elements present in the reflective process enables instructors to approach their students in a variety of ways to optimize their reflective abilities. In this study, the researchers strove to capture the breadth and depth of the elements of reflection. Unlike previous studies, which utilized the theories of Boud et al.[16] and Mezirow,[9] the coding schema developed in this method of assessment utilized each of the theoretical frameworks of Boud et al.,[16] Mezirow,[9] and Schön.[17]

Williams et al.[10] chose not to use the criteria from Mezirow,[9] stating that they preferred to focus on the process of reflection rather than on what they suggest to be categorization of the different types of reflection. However, the researchers in the current study believed that the elements of Mezirow's[9] reflective process enabled them to pinpoint the level of reflection present in each journal. This assessment can provide instructors with a baseline from which to more effectively and efficiently begin to facilitate the reflective process in their students. Further, incorporation of Mezirow's[9] elements of reflection provides an additional perspective upon which competence in the reflective process can be judged.

While Williams et al.[10] chose not to utilize Mezirow's[9] framework, Wong et al.[12] did use this framework to categorize students as nonreflectors, reflectors, or critical reflectors. Similarly, the current researchers saw value in utilizing Mezirow's[9] framework to assess the level of competence evident within a given journal; however, they chose to categorize journals rather than students. This decision was based on the fact that while journals provide instructors with insight into their students' thought processes, this insight is limited. Although there may be no evidence of reflection present in the particular journal entries submitted, it cannot be assumed that the student is a nonreflector; therefore, the researchers believed that they should not be judged in this manner.

While Wong et al.[12] used criteria from Boud et al.[16] and Mezirow[9] to assess reflective journals, the current researchers believed that the work of Schön[17] is equally vital to reflective practice, particularly in the health care environment, where practice is continually changing. Therefore, the current researchers chose to incorporate Schön's[17] criteria into the assessment schema presented.

Previous researchers have demonstrated reliability in assessing reflective journal writing while using different reflective criteria.[10,12,21] The results of this study support the fact that journals can be reliably assessed. While Wong et al.[12] showed reliable results (r = 0.88) using Mezirow's[9] work, they failed to achieve reliable outcomes with criteria developed from Boud et al.[16] The methodology used in the current study successfully minimized the difficulty encountered by Wong et al.[12] by looking more broadly at the three major elements of the framework of Boud et al.[16] (i.e., returns to experience, attends to feelings, evaluates the experience) rather than at the more detailed and recursive elements of integration, validation, association, and appropriation. Further, no attempt was made to match the quotes that were coded; rather, the researchers chose to look more broadly at evidence of the elements of reflection present within the overall journal.

Alternatively, Williams et al.[10,11] were successful in using the work of Boud et al.[16] by adapting their criteria. However, journal scoring was not clearly explicated in either study. Further, they created a linear scale from the work of Boud et al.[16] and noted that there was no evidence of improved levels of reflection over time. Yet, Boud et al.[16]

TABLE 4. Intraclass Correlation Coefficients, With 95% Confidence Intervals, Comparing All Three Raters on the Nine Elements of Reflection

Reflection Component	ICC(2,1)	95% CI
Reflection in action	0.55	0.37, 0.70
Reflection on action on	0.41	0.22, 0.59
Reflection for action	0.60	0.44, 0.74
Content	0.60	0.43, 0.74
Process	0.44	0.25, 0.61
Premise	0.72	0.59, 0.82
Returns to experience	0.03	–0.11, 0.23
Attends to feelings	0.28	0.10, 0.47
Reevaluates	0.43	0.21, 0.62

stated that reflection is not necessarily a linear process but rather is more recursive in nature; thus, a linear scale might not necessarily capture changes in a student's skill level. While the study by Williams et al.[11] yielded a score for each journal based on the criteria of Boud et al.,[16] the current researchers were interested in determining whether students evidenced use of the various elements of the reflective process in their writing rather than in providing an overall score or grade for each journal. The method of assessment developed in this study provides the instructor with greater information about student competence and can provide a baseline for facilitating the variety of elements of reflection needed to enhance both the breadth and the depth of the reflective process.

Finally, Pee et al.[21] developed a reliable measure, utilizing a structured worksheet, which looked at the depth of reflection but did not consider the breadth of the reflective process evident in student journals. As noted, by staying close to the theory, the current study provided a reliable method of assessment that incorporated both the breadth and the depth of the reflective process. Assessing the elements of the reflective process as described by Mezirow,[9] Boud et al.,[16] and Schön[17] provides breadth, while assessing the level of reflection as described by Mezirow[9] provides depth of the reflective process present in the journals.

Conclusions and Implications for Practice

This study presents a reliable method of assessing reflective journals. Staying close to the theory behind the reflective process provides a mechanism by which several elements from a number of theoretical frameworks on reflective practice can be used to judge reflective journal writing. It offers a method of determining whether, and to what level, students show evidence of reflection in their writing and as a result enables academic and clinical instructors to more efficiently and effectively determine where to begin facilitating the reflective process in their students. In addition, the coding schema developed provides a framework for teaching students about the various components of reflection critical to reflective thinking and reflective practice. Facilitating the different elements of reflection can be used to extend the reflective thought process and enhance the critical thinking skills of students. Finally, this method of assessment provides a mechanism for instructors to judge competence in the reflective process without judging the personal content or topic of the student's reflections. It separated the product from the process, allowing competence to be judged without undermining the freedom of thought critical to the reflective process.

Recommendations

The ability to analyze the elements of reflection is critical for effective assessment of reflective journals. Additional research is needed to determine the generalizability of this assessment method to other students and other raters in both academic and clinical settings. While this study provides a mechanism to assess whether, and to what level, students show evidence of reflection in their journals, further research is needed to determine whether facilitation of the elements of reflection can enhance reflective thinking and to determine the impact of reflective thinking on clinical practice. In addition, given that a significant percentage of the journals submitted showed evidence of reflection, qualitative analysis of the data would elicit reflective themes from the journals and allow instructors to better understand students' perceptions of their academic and clinical experiences.

Appendix

Operational Definitions for the Coding Schema

Level I Unit of Analysis: Words, Sentences, Paragraphs

I. Time dependent (Schön and Killion and Todnem)

Reflection in action (R-I-A). Occurs while in the midst of an activity and is the result of conscious decisions made on the spot.

Reflection on action (R-O-A). Occurs after the action has been completed. Reflection on action occurs only when the student (1) provides a description of a specific event and (2) attempts to better understand the situation, his or her action, and/or the outcomes.

Reflection for action (R-F-A). The student begins to anticipate situations before being faced with them and/or begins to plan for the future to improve the present situation/outcome.

II. Content dependent (Mezirow and Cranton)

Content (CON). The student attempts to explore the problem/experience to better understand it. The student goes beyond just describing an event to exploring the problem or situation. The student may begin to view the problem/experience from different perspectives.

Process (PROC). The student begins to describe the strategies and/or processes involved in an experience. These may

include strategies/processes used in learning, problem solving, or managing a situation. The student may begin to explore other strategies available for use.

Premise (PREM). The student recognizes and begins to explore or critique his or her own assumptions, values, beliefs, and biases. The student may begin to seek multiple perspectives and alternative explanations.

III. Stage dependent (Boud et al.)

Returns to experience (RETRN). The student describes an experience replaying what he or she considers significant.

Attends to feelings (ATTEND). The student acknowledges and begins to work with feelings that might have resulted from the experience.

Reevaluates (RE-EVAL). The student reappraises the current situation vis-à-vis past experiences. The student might (1) associate the experience with past experiences to link new concepts to preexisting knowledge; (2) begin to integrate new information; (3) validate his or her new insights, thoughts, and perceptions (i.e., check the authenticity), and (4) appropriate this new meaning into his or her own way of being.

Level II Unit of Analysis: the Journal (Jarvis et al., Mezirow, and Schön)

Nonreflection (NR). No evidence of reflection is present within the journal. The writer may describe experiences with no evidence of questioning or evaluation of the experience. Lack of reflection implies a person who acts based on habit or what he or she already knows, makes assumptions, acts mechanically, may not consider the potential for learning or change, and may even reject the possibility of learning something new because he or she is sure he or she is right.

Reflection (R). Evidence of reflection is present in the journal. This implies evidence that the writer either pauses in action or ex post facto to explore an experience, with the intent of better understanding the situation, or to decide how best to perform. This writer moves beyond simply reporting or describing events, to attempting to understand, question, or analyze the events.

Critical reflection (CR). Evidence of critical reflection is present within the journal. This implies evidence of a writer who stops to explore the existence of the problem, where the problem stems from, or the assumptions underlying the problem. The writer revisits the experience, begins to critique his or her own assumptions and thought processes, shows evidence of recognizing his or her own assumption, and may begin to show evidence of modifying his or her own biases or assumptions. This person typically shows evidence of premise reflection.

References

1. *Commission on Accreditation of Physical Therapy Education Accreditation Handbook*. Alexandria, VA: Department of Accreditation in Physical Therapy Education; 1997.
2. Boud D: Using journal writing to enhance reflective practice. *New Dir Adult Continuing Educ* 2001; 90:9–18.
3. Bourner T: Assessing reflective learning. *Educ Training* 2003; 45: 267–272.
4. Jarvis P: Reflective practice and nursing. *Nurse Educ Today* 1992; 12:174–181.
5. Jarvis P: Journal writing in higher education. *New Dir Adult Continuing Educ* 2001; 90:79–86.
6. Jensen G, Denton B: Teaching physical therapy students to reflect: a suggestion for clinical education. *J Phys Ther Educ* 1991; 5(1):33–38.
7. Kerka S: Journal writing and adult learning. *ERIC Digest No. 174*. 1996:1–4.
8. Kerka S: Journal writing as an adult. *ERIC: Learning Tool Practice Application Brief No. 22* 2002:84–89.
9. Mezirow JA: *Fostering Critical Reflection in Adulthood: A Guide to Transformative and Emancipatory Learning*. San Francisco, CA: Jossey-Bass; 1990.
10. Williams R, Sundelin G, Foster-Seargeant E, et al: Assessing the reliability of grading reflective journal writing. *J Phys Ther Educ* 2000; 14(2):23–26.
11. Williams RM, Wessel J, Gemus M, et al: Journal writing to promote reflection by physical therapy students during clinical placements. *Physiother Theory Pract* 2002; 18:5–15.
12. Wong FK, Kember D, Chung L, et al: Assessing the level of student reflection from reflective journals. *J Adv Nurs* 1995; 22:48–57.
13. Boenink AD, Oderwald AK, De Jonge P, et al: Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity, and initial experiences. *Med Educ* 2004; 38:368–377.
14. Williams RM, Wessel J: Reflective journal writing to obtain student feedback about their learning during the study of chronic musculoskeletal conditions. *J Allied Health* 2004; 33:17–23.
15. Boyd E, Fales A: Reflective learning: key to learning from experience. *J Hum Psychol* 1983; 23:99–117.
16. Boud D, Keogh R, Walker D: *Reflection: Turning Experience into Learning*. New York, NY: Kogan Page/Nichols; 1985.
17. Schön DA: *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass; 1987.
18. Killion J, Todnem G: A process for personal theory building. *Educ Leadership*. 1991; 48(6):14–16.
19. Atkins S, Murphy K: Reflection: a review of the literature. *J Adv Nurs* 1993; 18:1188–1192.
20. Shepard KF, Jensen GM (eds): *Handbook of Teaching for Physical Therapists*. Boston, MA: Butterworth-Heinemann; 1997.
21. Pee B, Woodman T, Fry H, et al: Appraising and assessing reflection in students' writing on a structure worksheet. *Med Educ* 2002; 36:575–585.
22. Kalliath T, Coghlan D: Developing reflective skills through journal writing in a OD course. *Organ Dev J* 2001; 19(4):61–70.
23. Shields E: Reflection and learning in student nurses. *Nurse Educ Today* 1995; 15:452–458.
24. Brookfield SD: *Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting*. San Francisco, CA: Jossey-Bass; 1987.
25. Mezirow J: *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass; 1991.
26. Schön DA: *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books, Inc; 1983.
27. Gandy J, Jensen G: Group work and reflective practicums in physical therapy education: Models for professional behavior development. *J Phys Ther Educ* 1992; 6:6–10.
28. Sugerman DA, Doherty KL, Garvey DE, et al: *Reflective Learning: Theory and Practice*. Dubuque, IA: Kendall/Hunt; 2000.
29. Trysennar J: Interactive journals: an educational strategy to promote reflection. *Am J Occup Ther* 1995; 49:695–702.
30. Woodward H: Reflective journals and portfolios: learning through assessment. *Assess Eval Higher Educ* 1998; 23:415–423.
31. Brookfield SD: *Becoming a Critically Reflective Teacher*. San Francisco, CA: Jossey-Bass; 1995.
32. Cranton P: *Understanding and Promoting Transformative Learning: A Guide for Educators of Adults*. San Francisco, CA: Jossey-Bass; 1994.

33. Gordon J: Assessing students' personal and professional development using portfolios and interviews. *Med Educ* 2003; 37:335–340.
34. Kennison MM, Misselwitz S: Evaluating reflective writing for appropriateness, fairness, and consistency. *Nurs Educ Perspect* 2002; 23:238–242.
35. Drevdahl D, Dorcy KS: Using journals for community health students engaged in group work. *Nurse Educ* 2002; 27:255–259.
36. Jarvis P, Holford J, Griffin C: *The Theory and Practice of Learning*. London, England: Kogan Page; 1998.
37. Domholdt E: *Physical Therapy Research: Principles and Applications*, 2nd ed. Philadelphia, PA: Saunders; 2000.
38. Portney LG, Watkins MP: *Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Prentice Hall; 2000.
39. Miles AM, Huberman MB: *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed. Thousand Oaks, CA: Sage Publications; 1994.
40. Fleiss JI, Cohen J: The equivalence of weighted kappa on the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; 33:613–619.
41. Rae G: On measuring agreement among several judges on the presence or absence of a trait. *Educ Psychol Meas* 1984; 44:247–253.
42. Siegel S, Castellan NJ Jr: *Nonparametric Statistics for the Behavioral Sciences*. New York, NY: McGraw Hill; 1988.
43. Kerlinger FN, Lee HB: *Foundations of Behavioral Research*, 4th ed. Philadelphia, PA: Harcourt College Publishers; 2000.
44. Gronlund NE: *Measurement and Evaluation in Teaching*, 5th ed. New York, NY: Macmillan; 1985.
45. Johnson RL, McDaniel F II, Willeke MJ: Using portfolios in program evaluation: an investigation of interrater reliability. *Am J Eval* 2000; 21:65–80.
46. Rea LM, Parker RA: *Designing and Conducting Survey Research: A Comprehensive Guide*, 2nd ed. San Francisco, CA: Jossey-Bass; 1997.