International Journal of ePortfolio
http://www.theijep.com

2012, Volume 2, Number 1, 75-85
ISSN: 2157-622X

Analysis of a Rubric for Assessing Depth of Classroom Reflections

Dev K. Dalal, Milton D. Hakel, Michael T. Sliter, and Sarah R. Kirkendall
Bowling Green State University

Writing reflections is recommended for enhancing retention and transfer of learned material. The benefits of student reflections have been well documented, but the methods for collecting and assessing reflections can be difficult. This study presents the development and analysis of a new, straightforward rubric for assessing depth of student reflections. The psychometric properties of the depth ratings based on the rubric and preliminary validity evidence of the ratings are investigated. With this rubric, raters were able to assess the depth of reflections very reliably. Depth ratings were significantly related to GPA.

Retention and transfer of learned material are important—although too often merely implicit—goals of classes at the university level. Factors that enhance retention and transfer of learning are now key areas of research, and one frequently recommended method of enhancing these goals is to incorporate reflection into classroom practices (Saito & Miwa, 2007). Reflection is the ability to think critically about successes and failures, extract ideas and information from a variety of sources, and recognize when current information can be used in the future (Hopkins, 1997). Reflections and reflective ability have been linked to many positive academic outcomes including retention and transfer of learned material (Cassidy, 2006). However, there is a substantial gap between the findings from research on reflection and their application. In particular, instructors who wish to incorporate reflections into their classes may struggle to find a method that is tractable yet effective for both collecting reflections and assessing their depth. This report offers a model designed to facilitate the collection and assessment of student reflections. First, we outline how electronic portfolio (e-Portfolio) systems offer simple methods for collecting reflections in the classroom. Collecting and managing reflections from a large group of students can be logistically difficult; however, we demonstrate how e-Portfolio systems offer an approach to deal with this challenge. Second, we present a rubric that we believe will allow for peer assessments of reflection depth. Indeed, assessing the reflections of a large group of students can be a daunting task for an instructor. Here, we present a rubric and training session that can allow for reliable peer assessments of the depth of the reflections taking some of the burden off the instructor. We also present an assessment of the reliability of these depth ratings and some evidence of the validity of these assessments. Results of these analyses will demonstrate that this rubric is a useful tool for peer assessments of student reflections as collected via e-Portfolios.

As stated above, reflection is the ability to think critically about successes and failures, extract ideas and information from a variety of sources, and recognize when current information can be used in the future (Hopkins, 1997). More simply, reflecting is engaging in meta-cognitive activities to assess one's learning. As such, reflection is seen as similar to other cognitive activities including meta-cognition (Saito & Miwa, 2007) and self-assessment (Cassidy, 2006). These three concepts share the idea of evaluating one's learning and using the evaluations to further understand and apply newly learned materials.

King and Kitchener (2004) developed a model of reflective judgment formulated around the concept of epistemic cognition—underlying assumptions about what knowledge is and how it is gained. They further demonstrate that reflective thinking develops slowly over time from adolescence to adulthood as one's thinking begins to recognize the uncertainty in knowledge. This work on reflective thinking underscores the point that reflective ability develops over time and may need formal development.

The benefits of reflection have been outlined in many studies. In particular, Saito and Miwa (2007) demonstrated that reflecting aided in retention of learned material. Furthermore, Cassidy (2006) showed that reflection was related to deeper learning. Boyle, Duffy, and Dunleavy (2003) demonstrated that deep learners are characterized by being intrinsically motivated, able to relate new materials to previously learned information, and able to critically evaluate information; deep learning is also positively related to grade point average (GPA) and average exam scores.

An interesting application of reflection in an educational setting was reviewed by Knowles, Borrie, and Telfer (2005). These researchers showed that reflection in a sport coaching program helped develop more effective coaching, noting that those coaches who are taught to reflect on both their performance as a coach and the performance of the team were able to coach more effectively. They go on to advocate for the implementation of reflection training into all elite coaching programs. Finally, reflection has been linked

to deeper thinking (Short & Rinehart, 1993). Journal entry reflections were quantitatively and qualitatively analyzed with the results showing that levels of reflection and complexity of thinking increased by the end of the year. In all, these studies demonstrate that reflecting is associated with beneficial outcomes.

A common feature in many studies of reflection is the complicated method for scoring reflections. For example, researchers tend to use highly structured methods for gathering reflections; these methods include semi-structured interviews (King & Kitchener, 2004; van Kraayenoord & Paris, 1997) or a structured diary paradigm (Short & Rinehart, 1993). Furthermore, reliable scoring of reflections using these methods typically requires highly trained coders (King & Kitchener, 2004). Although these methods are appropriate for the research on reflections in the classroom, their complexity poses substantial constraints when trying to implement and assess reflections for large numbers of students. What is needed to make it practical, then, is a relatively straightforward medium for collecting reflections and an easy method for evaluating those reflections. e-Portfolio systems provide a medium to address this first issue.

e-Portfolios

The use of student portfolios as an assessment tool has increased since the 1990s (Ewell, 2002). Furthermore, portfolios provide opportunities to collect feedback to improve curricula and student performance (Ewell, 2002). e-Portfolios offer the benefits of paper and pencil portfolios while adding the ability to highlight audio-visual artifacts, to be viewed from remote locations, and to reflect on portfolio artifacts (Cambridge, 2001). The versatility of e-Portfolios presents instructors with a powerful method for gathering assessment information for their students (Goldsmith, 2007). Moreover, because e-Portfolios are Internet based, instructors and peers can access e-Portfolios at any time allowing for assessments outside of the classroom setting. These features of e-Portfolios suggest that they offer a medium to collect and maintain a large number of student reflections.

e-Portfolios have the versatility to house both global assessments and evaluations of specific facets of student learning. In addition, e-Portfolios offer diverse opportunities for enhancing teaching and learning. Research on e-Portfolios has uncovered positive relationships between e-Portfolio usage and various academic outcomes. Knight, Hakel, and Gromko (2008) showed that e-Portfolio users had higher overall GPAs, credit hours earned, and rates of retention in college. Furthermore, they showed that GPA and credit hours earned were positively related to the number of artifacts uploaded and number of versions of resumes uploaded. One challenge in integrating reflections into the classroom is the difficulty with collecting reflections; e-Portfolios represent one medium for collecting reflections.

Yancey (2009) demonstrated how e-Portfolio systems might be used to collect reflections from students. She further demonstrated that the structure of the e-Portfolio system might have an impact on student reflection. More specifically, Yancey reviewed the e-Portfolio systems of various universities and determined that the system design can impact whether or not students will reflect. e-Portfolios are a medium to encourage and catalog reflections that can be viewed from anywhere. Yancey further showed that students who reflected (as opposed to those who did not reflect) were more engaged and reported more benefits of learning.

Tigelaar, Dolmans, De Grave, Wolfhagen, and van der Vleuten (2006) investigated the role of creating and maintaining portfolios to stimulate teacher reflections, finding that assignments to update portfolios lead to increases in the frequency of reflecting about teaching. In particular, reflections within the portfolio were centered on teaching effectiveness and functioning. Similarly, Groom and Maunonen-Eskelinen (2006) showed that maintaining a portfolio helped student teachers engage in reflections and develop the ability to self-assess their teaching.

The studies outlined above demonstrate how portfolios can stimulate and collect reflections, and how reflecting can improve engagement and learning relative to the absence of reflection. These studies do not, however, investigate the depth of these reflections. Indeed, none of the above studies assessed the depth of these reflections, nor attempted to assess the relationship between depth and academic outcomes. Although reflecting is better than not reflecting (see Yancey, 2009), it is important to assess the depth of these reflections, ascertain whether the reflection depth is related to student outcomes, and determine whether or not instructors should look to encourage deeper reflection. Interestingly, no studies investigating the depth of student e-Portfolio reflections could be located—this study hopes to address this gap in the literature.

Reflections and e-Portfolios

Independently, research on depth of reflections and e-Portfolios has shown many benefits; however, there is little research investigating the depth of e-Portfolio reflections. One primary reason for this is likely due to the lack of a straightforward method for assessing the depth of these reflections. Here, we look to help address this gap in the literature by providing a rubric to assess

reflection depth that can be integrated into an e-Portfolio platform. With this in place, then, it is possible to expand on the research investigating the role e-Portfolios can play in developing reflective ability in students and enhancing student learning.

Even though recent work has demonstrated how e-Portfolios can be used to encourage and collect reflections (see Yancey, 2009; Rickards & Guilbault, 2009), there are still more ways to utilize e-Portfolio reflections to enhance student learning. One example is to use e-Portfolio reflections to address some principles of learning. Pashler et al. (2007) and Graesser, Halpern, and Hakel (2008) identified principles of learning to help guide research and practice in education. Reflecting on e-Portfolio artifacts can help address these principles to improve student learning. Pashler et al. (2007), for example, suggested that instructors should encourage students to connect and integrate abstract and concrete representations of concepts. Reflecting on actual e-Portfolio artifacts can assist students in connecting the concrete artifact to the abstract lessons that the instructor was trying to convey.

The principles outlined by Graesser et al. (2008; see also Graesser, 2009) can similarly be integrated using reflections. For example, "explanation effects" (Principle 17) suggest that student learning is enhanced from constructing deep coherent explanations of the material; this is similar to deep thinking—as stated above, reflective ability has been linked to deep thinking and deeper reflections can stimulate deeper thinking. As another example, "deep questioning" (Principle 18) is an activity to help students learn better by asking questions that require the students to critically evaluate what they have learned. Reflection exercises can be responses to deep questions. As a final example, reflections can help "anchor learning" (Principle 25). When material is connected to real-world problems or applications, learning is deeper; reflections that connect an artifact to practical applications can help anchor the learned materials. Instructors can use e-Portfolios to encourage reflections to address these principles of learning.

The benefit of combining reflections with e-Portfolio artifacts allows students to gain a deeper appreciation for assignments and lessons. As stated above, deeper reflection is associated with deeper learning. Along these lines, then, it is important to begin to assess the depth of e-Portfolio reflections. We present a reasonably straightforward method for evaluating the depth of these reflections to accomplish this goal to expand the research on e-Portfolio reflections.

The Current Study

The current study presents a rubric for assessing the depth of e-Portfolio reflections. As such, presented first is a description of the rubric created for assessing reflection depth; this is followed by an examination of its psychometric properties. The second part of the study will present an investigation of whether depth of reflections is related to academic outcomes. This test provides preliminary validity evidence of the depth ratings obtained by using the depth rubric. To this end, we present two research questions to be investigated in the study:

- *Research Question 1:* How reliably can reflection depth be rated?
- *Research Question 2:* Do the depth ratings obtained using the depth rubric correlate with academic outcomes?

Methods

Participants

To answer question 1, nine raters were asked to rate the depth of a series of reflections. Three of the coders were graduate students in psychology, four were undergraduate psychology students, one was a non-student research assistant, and one was a faculty member in the psychology department.

The reflections focused on an uploaded file, or artifact, copied from the Epsilen e-Portfolio system from Bowling Green State University (BGSU). Hakel and Smith (2009) provide a description of the e-Portfolio system at BGSU. A total of 1,456 reflections from 324 different e-Portfolio users were collected. There was no consistency in the uploaded reflections. Some of the reflections were from specific classes whereas others were just unprompted reflections uploaded by the user. All of the reflections, however, were unstructured and largely unguided. Because of this, the depths of these reflections are likely to be rather inconsistent.

Reflections from 219 (note: sample sizes of actual analyses varied from 219 due to missing data) students from the same university were gathered to investigate question 2. All of the students had undergraduate credits while 35 students (16%) had some graduate credits. These 35 students were undergraduate students who took graduate classes, undergraduates who continued into graduate school, or graduate students who took undergraduate courses. These students were retained in the analyses since their undergraduate outcome variables were available for the analyses. Demographically, 63% of the sample was female and 88% of the sample reported ethnicity as white. Five percent of the sample consisted of freshmen, 2.7% sophomores, 3.6% juniors, and 72.7% of the sample consisted of seniors; the remaining respondents (~16%) listed their class level as "other."

Materials and Procedures

Depth rubric and coding. Reflections were independently coded by the nine member research team using the Reflections Depth Rubric created for the study (Appendix A). The rubric consists of six ratings of depth, ranging from 0 meaning *not a reflection* (e.g., "test upload") to 5 meaning *deep reflection*. Recall from above that reflection is the ability to think critically, extract ideas, and apply current information to future uses (Hopkins, 1997). Using this definition, the rubric was developed in the view that reflections become deeper to the extent that the reflector extrapolates from the actual artifact or assignment to higher level applications and deeper thinking. That is, deeper reflections will center less on the actual uploaded artifact, and will demonstrate a critical evaluation of the learning behind the artifact, and how the principles are related to future application. For example, note this student's reflection: "This file demonstrates my skills using Excel, and how I used Excel in an M&M lesson to incorporate technology with my students." Based on the rubric developed, this reflection received an average rating of 0.81 because the student did not reflect past the actual artifact to demonstrate an understanding of the applications of the Excel file past the immediate purpose, nor did the student critically evaluate the learning that occurred while developing the artifact. Take, instead, this excerpt from a student's reflection:

> When writing the paper, I spent a great deal of time critically thinking about what I feel as a person compared to my beliefs in the professional field. The process involved me learning more about myself and integrating different aspects of my life into my professional field. . . . From this assignment, I feel that I have grown as an FLE because I have identified my own values and feelings that impact me as a professional in FLE.

The whole reflection had an average score of 4.52 since this reflection demonstrates reasoning past the immediate purpose of the artifact; indeed, this student was able to anchor the artifact to future uses. Note that this rubric defines reflections to be deeper the more the student demonstrates that he/she has critically evaluated the learning behind and development of the artifact, and to the extent that he/she understands that the artifact exists beyond the confines of the assignment and begins to evaluate what was learned or the artifact itself can inform future work—in short, deeper reflections represent deeper and more critical thinking about learning which have been shown to be positively related to academic outcomes (see above, Appendix A provides further examples of reflections of different depths).

After a short training sessions and practice codings (see Appendix B), the coders assessed a total of 1,456 reflections from 324 different e-Portfolio users. Coders were asked to make a primary depth rating, and also to make a secondary depth rating if the reflection did not fit one of the defined cut points exactly. For example, if a reflection was between a 2 and 3 in terms of depth, the coder was allowed to use the primary rating option for the most representative depth, but recorded a second rating option for the other (e.g., primary of 2, secondary of 3). Final reflection depth scores were calculated by computing a weighted average where the primary rating received a weight of 2/3 and the secondary rating received a weight of 1/3. This weighting scheme was used to give more emphasis to the coders' initial thoughts while still allowing for uncertainty.

Dependent measures. The dependent measures for research question 2 were obtained by matching university ID numbers with data from the Institutional Research office of the university. These analyses used Cumulative GPA, total credit hours earned, and ACT scores as dependent measures of academic outcomes. Using a concordance table from Doran, Lyu, Pommerich, and Houston (1997), individuals' with only SAT scores had their scores converted to ACT scores.

Results

Research Question 1

To answer question 1, a two-way random effects intraclass correlation (ICC) was computed. This ICC assumes that both the reflections and the raters are random samples of their respective populations. We make the assumption of random effects of the raters as we expect the raters for this study to represent potential evaluators of artifact reflections. Inter-rater consistency was measured as we were not interested in absolute agreement between the weighted average depth scores, but were instead interested in consistency of ranking of the reflections (i.e., across all raters, reflection 100 is deeper than reflection 102). Therefore, an ICC(2,1) was calculated for the nine judges over all the reflections (McGraw & Wong, 1996; Shrout & Fleiss, 1979).

The ICC for averaged measures based on the nine raters over the 1,456 reflections was 0.946 (F (1455, 11640) = 18.47; $p < .001$; 95% CI = 0.942 to 0.950). This value is well above the recommended ICC = 0.70 value before aggregating data (e.g., Klein & Kozlowski, 2000). Furthermore, the lower bound of the confidence interval is well above this criterion. The ICC for averaged measures was used since the average weighted depth rating of the coders is the most appropriate index of reflection depth. The ICC value for just the four undergraduate students was significantly lower than nine raters (ICC(2,1) = .90, F (1455,4365) =

Table 1
Means, Standard Deviations, and Intercorrelations of Deptartment Ratings and Academic Outcome Variables

	M	*SD*	1	2	3	4
1. Maximum Depth	1.61	1.15	1.00			
2. ACT	21.69	4.00	0.05	1.00		
3. GPA	3.17	0.53	0.27**	0.51**	1.00	
4. Credit Hours Earned	116.50	36.96	0.10	0.03	0.13	1.00

Note. Ns = 170 – 219; ** Correlation is significant at the 0.01 level (2-tailed).

9.96, $p < .001$, 95% CI = 0.89 – 0.91). However, this is expected given the reduced number of raters; also, the ICC value is still well above the recommended cut-off.

Here, we used the average rating of nine raters to determine the level of reliability of the ratings. However, using nine raters per reflection is wasteful if fewer raters can be used while still obtaining the desired level of reliability. An equation can be used to estimate the number of raters needed to attain a certain level of reliability or ICC value:

$$k = \frac{ICC^*\left(1 - rl\right)}{rl\left(1 - ICC^*\right)} \qquad (1)$$

Where k is the number of raters, ICC^* is the desired ICC value, and rl is the lower bound of the 95% confidence interval for the estimated ICC value. Setting ICC^* = 0.946, k = 1.08 indicating that a minimum of two raters can be used to obtain an ICC value of 0.946.

Research Question 2

Each reflection's final depth rating was an average of the nine coders' weighted depth scores. Some students uploaded more than one reflection; in fact, the number of reflections uploaded ranged from one to 44. For this analysis, only the students' maximum reflection depth was used. For those students who wrote only one reflection, this one reflection was treated as their maximum depth. Depth assessments were correlated with the dependent measures to establish the relationship between depth of reflection and academic outcomes.

Table 1 presents the correlations between the variables as well as the means and standard deviations of the measures. As the table shows, the mean of the maximum depth rating was just above 1 (M = 1.61, SD = 1.15). The mean value suggests that, even at their deepest reflection, these students are not reflecting at a very deep level. This is to be expected since these reflections were free response without any guidance. It is unlikely that students would reflect at a very deep level without some form of instruction. Indeed, King and Kitchener (2004) demonstrated the reflective ability develops overtime, and Yancey (2009) showed that the structure of e-Portfolio system can impact reflections. Maximum reflection depth was significantly related only to GPA. Reflection depth, as judged by the rubric, was unrelated to credit hours earned and ACT scores.

Discussion

The results of this study showed that raters can be trained to rate reflection depth reliably. The simplicity of the rubric and training makes this method for assessing reflection depth a viable alternative to the more demanding approaches used in previous research (for examples see King & Kitchener, 2004; Short & Rinehart, 1993; van Kraayenoord & Paris, 1997). Furthermore, the ICC values are well above the cut-off recommended for aggregating scores, showing that ratings can be aggregated to create a final depth score of a student's reflection. As equation 1 showed, an ICC value as large as found in this study could be found with as few as two raters. In answer to question 1, therefore, this rubric in combination with a short training session can result in reliable assessments of reflection depth. This suggests that instructors can create triads of students where two students rate the third student's reflections with a strong level of rater reliability.

Research question 2 asked about preliminary validity evidence of the ratings. Although the maximum weighted depth ratings were unrelated to credit hours earned and ACT scores, initial validity is demonstrated by the moderately positive relationship between the ratings and GPA. These results make sense, however, when one considers the temporal ordering of the variables. The ACT is taken before the student begins to reflect in college. Research has shown that reflective ability becomes better and reflections become deeper with time (King & Kitchener, 2004), and even without instruction students would deepen their reflections even a little over time. Therefore, reflective ability in college would not necessarily be related to ACT scores from high school. Therefore, in response to question 2, this study has shown preliminary evidence of the validity of these ratings.

Table 2
Examples of Reflections Addressing Learning Principles

Principle Number[1]	Description of Principle[1]	Reflection Text[2]
17	**Explanation Effects**. Students benefit more from constructing deep coherent explanations.	This lab was one of my favorites! I am very interested in manipulating foods, especially in baking. Sweeteners are something I never really tampered with becuase I did not want to ruin my product, but after this lab I learned so much and have a better understanding of sugar and substitutes now. Sugar gives the beautiful browning when foods are baked which makes it look appetizing, plus, the structure of sugar cuts air into the product to give it extra rise. A full 100% substitution for sugar is not a good idea; the product is flat and does not look nearly as appetizing. This lab gave me knowledge that I can apply when I bake at home.
18	**Deep questions**. Students benefit more from asking and answering deep questions.	The learning outcome that was demonstrated in this assignment was inquiry. We were asked what were valuable assets that a scientist should have. This assignment was very useful in the fact that it required us to think about what would make a good scientist and what a good scientist would have to embody. This paper required a lot of thought and conversation between both Stephanie and I because this was very thought provoking. Some of the ideas that we had come up between the both of us had come up twice and that was when we knew that that particular trait/characteristic was important.
25	**Anchored Learning**. Learning is deeper and students are more motivated when the materials and skills are anchored in real world problems that matter to the learner.	My internet scavenger hunt will be used directly with my social studies Caribbean cultures unit. The scavenger hunt will allow for my students to learn about different countries flags (something they love to learn about) and a quick fact about each Caribbean nation selected. As a teacher, I can use this as a formal assessment of their internet, word processing, and research skills. I would like to create at least one internet scavenger hunt for each of my major teaching units.

1. From Graesser, Halpern, & Hakel (2008)
2. Note: Errors in text represent typographical errors in original reflection text made by the student.

As described above, incorporating reflections into the classroom is an important opportunity to increase depth of learning in the classroom. This study adds to this growing body of literature by presenting a new method for collecting and assessing reflections to make this incorporation more feasible. Past research has shown that students with active e-Portfolios have higher GPAs, credit hours earned, and retention rates than a matched sample of non-users (Knight et al., 2008). By incorporating the use of e-Portfolios into the classroom, the benefits to students will accrue first from keeping an updated e-Portfolio (Knight et al., 2008), but also by fostering reflections (Yancey, 2009).

e-Portfolios are a good medium to collect reflections. The benefits of reflection are likely due to the ability to critically evaluate the learned information and assist students in actively learning the information rather than relying on rote memorization. Additionally, the learning principles outlined by Graesser et al. (2008; e.g., Anchoring, Questioning) can be addressed by this technology because it offers the student a forum in which to reflect. Indeed, as Table 2 shows, even though these reflections were free recall and unguided, some of these reflections exhibit the learning principles outlined. e-Portfolios offer students the ability to catalog their work and critically evaluate that work.

This study provides a new approach for incorporating reflections into the classroom. By utilizing e-Portfolio systems and the rubric described here, instructors can make use of reflections in their classes. However, future research is needed to replicate and extend these results. Future research should look at more than just GPA as an outcome measure. Retention and transfer of learning are two variables related to knowledge, and their relationship to reflections should be studied explicitly rather than inferred from correlates (e.g., the relationship between reflection depth and GPA). Also, the reflection literature may be bolstered with an experimental investigation that randomly assigned students into either a reflection or non-reflection condition to see if the outcomes are better for those who reflected. Finally, future research can try to explicitly study the link between reflections and the principles of learning rather than inferring it. Even though many of the reflections in our sample demonstrate these learning principles (see Table 2), it is worth investigating the extent to which reflections structured around these principles improve learning. For example, one can study the role that reflection plays in being able to effectively anchor lessons to actual future application.

With this rubric and a versatile course management system like e-Portfolios, instructors can enhance student learning by incorporating reflections into their classes. Indeed, the rubric is offered as a tool that might be embedded into e-Portfolio, learning management, and course management software, a step that can benefit both students and researchers.

References

Boyle, E. A., Duffy, T., & Dunleavy, K. (2003). Learning styles and academic outcome: The validity and utility of Vermunt's inventory of learning styles in a British higher education setting. *British Journal of Educational Psychology, 73*(2), 267-290. doi:10.1348/00070990360626976

Cambridge, B. (2001). Electronic portfolios as knowledge builders. In B. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.), *Electronic portfolios: Emerging practices in student, faculty, and institutional learning* (pp. 1-11). Washington, DC: American Association for Higher Education.

Cassidy, S. (2006). Learning style and student self-assessment skill. *Education + Training, 48*(2/3), 170-177. doi:10.1108/00400910610651791

Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College and University, 73*, 24-31.

Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T. W. Banta and Associates (Eds.), *Building a scholarship of assessment* (pp. 3-25). San Francisco, CA: Jossey-Bass.

Goldsmith, D. J. (2007). Enhancing learning and assessment through e-Portfolios: A collaborative effort in Connecticut. *New Directions for Student Services, 119*(31), 31-42. doi:10.1002/ss.247

Graesser, A. C. (2009). Inaugural editorial of journal of educational psychology. *Journal of Educational Psychology, 101*(2), 259-261. doi:10.1037/a0014883

Graesser, A. C., Halpern, D. F., & Hakel, M. D. (2008). *25 principles of learning.* Washington, DC: Task For on Lifelong Learning at Work and at Home. Retrieved from http://www.psyc.memphis.edu/learning/whatweknow/index.shtml.

Groom, B., & Maunonen-Eskelinen, I. (2006). The use of portfolios to develop reflective practice in teacher training: A comparative and collaborative approach between two teacher training providers in the UK and Finland. *Teaching in Higher Education, 11*(3), 291-300. doi:10.1080/13562510600680632

Hakel, M. D., & Smith, E. N. (2009). Documenting the outcomes of learning. In D. Cambridge, B. Cambridge, & K. Yancey (Eds.), *Electronic portfolios 2.0: Emergent research on implementation and impact.* Sterling, VA: Stylus.

Hopkins, D. (1997). Improving the quality of teaching and learning. *Support for Learning, 12*(4), 162-165. doi:10.1111/1467-9604.00038

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist, 39*(1), 5-18. doi:10.1207/s15326985ep3901_2

Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*(3), 211-236. doi:10.1177/109442810033001

Knight, W. E., Hakel, M. D., & Gromko, M. (2008). The relationship between electronic portfolio participation and student success. *AIR Professional File*.

Knowles, Z., Borrie, A., & Telfer, H. (2005). Towards the reflective sports coach: Issues of context, education and application. *Ergonomics, 48*(11-14), 1711-1720. doi:10.1080/00140130500101288

Mason, R., Pegler, C., & Weller, M. (2004). e-Portfolios: An assessment tool for online courses. *British Journal of Educational Technology, 35*(6), 717-727. doi:10.1111/j.1467-8535.2004.00429.x

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46. doi:10.1037/1082-989X.1.1.30

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalf. J. (2007). *Organizing instruction and study to improve student learning*. Washington, DC: National Center for Education Research, Institute of Educational Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf.

Saito, H., & Miwa, K. (2005). Construction of a learning environment supporting learners' reflection: A case of information seeking on the Web. *Computers & Education, 49*, 214-229.

Short, P. M., & Rinehart, J. S. (1993). Reflection as a means of developing expertise. *Educational Administration Quarterly, 29*(4), 501-521. doi:10.1177/0013161X93029004006

Shrout, P. E., & Fleiss, J. L. (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428. doi:10.1037/0033-2909.86.2.420

Tigelaar, D. E. H., Dolmans, D. H. J. M., De Grave, W. S., Wolfhagen, I. H. A. P., van der Vleuten, C. P. M. (2006). Portfolio as a tool to stimulate teachers' reflections. *Medical Teacher, 28*(3), 277-282. doi:10.1080/01421590600607013

van Kraayenoord, C. E., & Paris, S. G. (1997). Austrian students' self-appraisal of their work samples and academic progress. *The Elementary School Journal, 97*(5), 523-537. doi:10.1086/461879

Yancey, K. B. (2009). Reflection and electronic portfolios: Inventing the self and reinventing the

university. In D. Cambridge, B. Cambridge, & K. Yancey (Eds), *Electronic portfolios 2.0: Emergent research on implementation and impact.* Sterling, VA: Stylus.

DEV K. DALAL received his M.A. in Applied Social Psychology from Loyola University, Chicago, and is a fifth year doctoral candidate in the Industrial-Organizational Psychology program at Bowling Green State University. His research interests include statistical and methodological issues in research and practice, measurement theory and application, and judgment and decision making.

MILTON D. HAKEL is an Ohio Eminent Scholar and Professor Emeritus at Bowling Green State University and founding president of the Alliance for Organizational Psychology, an international federation. His electronic portfolio is online at www.epsilen.com/mhakel.

MICHAEL T. SLITER, M.A. is a fifth year graduate student in the Industrial-Organizational Psychology program at Bowling Green State University. His primary research interests are in the areas of customer service, workplace mistreatment, counterproductive behaviors, workplace attitudes, and emotional labor.

SARAH R. KIRKENDALL is a fourth year graduate student in the Industrial-Organizational Psychology program at Bowling Green State University. Her research interests include judgment and decision making as well as employee selection.

Appendix A
Reflections Depth Rubric

The purpose of this rubric is to outline the process by which written document reflections are numerically coded. The coding scheme is done using a six point scale ranging from surface to deep. Below is the coding scheme and how to differentiate between the codes. (Please note that all spelling errors in the examples have been noted, but all quotes are in their original form).

Code of 0
• Not a reflection • "lkj;asdkjfg ai8ujrgm;qawerg"
Code of 1
• Represents a surface reflection • The reflection **centers around the artifact** uploaded and does not extend pass the particular artifact. • Example 1: "I made this graph in Excel. It is my all time favorite graph. I'me sure you can see how hard I worked on it." • Example 2: "I absolutely loved writing this paper because it forced me to dig deep into this issue and write my personal opinion, which at times, can be very strong. I think that it has some good points and is a worthwhile document."
Code of 2
• Represents a surface-moderately deep reflection • The reflection still centers around the artifact uploaded, but starts to extend past the particular artifact to include a discussion of the class or thoughts about working on the particular artifact. It **does not extend to anything past the artifact** though; all thoughts are centered on the artifact still. • Example 1: "I believe this paper displays my ability to research and write a thorough description of complicated processes. It also displays my understanding and mastery of the following National Science Education Standard: The earth is a system containing essentially a fixed amount of each stable chemical atom or element. Each element can exist in several different chemical reservoirs. Each element on earth can move among reservoirs in the solid earth, oceans, atmosphere, and organisms as part of biogeochemical cycles." • Example 2: "This document taught me how to use a basic tool such as draw to create images and pictures using technology. I learned how to manipulate the mouse and object on the computer to come up with a final product that looks such as this."
Code of 3
• Represents a moderate reflection • The reflections starts to extend past the artifact itself and starts to talk about how the creating the artifact has helped them for the **near future**. There are notions of growth in education, but not connected to the distant future. • Example 1: "This is my Teaching Project I for MUED 240: Introductory Music Field Experience. I liked doing this project because it gave me a chance to become more skilled in using technology. Also, I was able to identify good and bad concepts in my teaching." • Example 2: "I liked doing this because it forced me to think about what good webpages entail but I do not think that anyone else will find it that valuable because it was done by a young college student and may not be as reliable as other website checklists."
Code of 4
• Represents a moderate to deep reflection • The reflector makes only a loose connection to the actual artifact, and starts to look at how the project has helped them for the **distant future**. The reflection is less concerned with the particular events of the reflection, but now looks at how the process was helpful and informative for other things. • Example 1: "This lesson plan I saw in my observations and I really liked how he did this lesson. I think that it will help the students get interested in authentic german songs, as well as learning about grammar and new vocabulary." • Example 2: "I really liked doing this in class because I feel that it is important to recognize children's acheivements and hard work with something that they are able to hang on their parent's refrigerator."

Code of 5

- Represents a deep reflection
- The artifact is hardly mentioned, but the process of creating it and the uses are the topics of the reflection. The reflector is discussing how the process is going to help him/her obtain future goals. The reflector is able to see how the individual artifact will actually be able to help **areas in their future**.
- Example 1: "With all the changes that occurred with my assistantship institution, I was expected to go above and beyond my normal duties. One such responsibility was to complete a directors report highlighting all the events and activities that student activities did. Doing this allowed me to look more critically at what types of programs this office produces. Further, I can critically analyze what methods and steps I can take to produce better events in the future."
- Example 2: "After taking EFTL 302, I have learned a great deal of information that will be VERY beneficial in my future as a teacher! Although it was very stressful at times, I learned how important it is to incorporate technology nto my classroom. I hope that the school I get hired at will have computers in my classroom and will be willing to supply me with the technology I need. I truely hope to use it as much as possible, especially to allow students to use it for their own work."

Appendix B
Reflection Depth Ratings: Training Session Information

Goals

The training for rating reflection depth was developed with two goals in mind. First, the training was meant to maximize rater consistency by allowing raters to familiarize themselves with the task, practice rating reflections as a group, and to practice rating reflections individually. Second, the training was meant to be replicable, meaning that every effort was made to document the process to facilitate future efforts in rating reflection depth. The same general training program could be adapted to rate reflections using a different medium, and can be applied to large scale training.

Throughout the training exercises, raters were exposed to a variety of example reflections. All examples used throughout this training were chosen from a pool of actual e-Portfolio reflections. Reflections were chosen either to (a) fit well within the definitions of the different depth categories, or (b) to be somewhat ambiguous as to the depth category. This was meant to facilitate discussion and allow raters to calibrate on reflections that do not necessarily fit into any definitional category.

Definitions and Examples

In both PowerPoint and handout formats, all raters were provided with definitions of each level of depth, ranging from "0" (not a reflection) to "5" (deepest level of reflection). Along with each definition, one or more example reflections were given, providing the raters with concrete examples to facilitate calibration. Raters were allowed (and expected) to refer to these definitions examples throughout all other training exercises, as well as during the actual rating task. The depth rating of these examples was previously agreed upon by the primary researchers.

Group Rating Activity

Following discussion of the definitions and the examples, a discussion exercise took place. Raters were given several example reflections, one at a time, and asked to rate these based on the definitions and the examples. The raters were to write down their ratings, and then discuss why they chose these ratings. Whenever there was disagreement, the raters were asked to discuss their differing viewpoints, and to settle on a score. This provided raters a concrete opportunity to calibrate. The trainers provided their own ratings of the exercise based on their own knowledge, which provided further discussion if there was disagreement.

Individual Rating Activity

To complete the training, each rater was given an individual rating activity. This activity presented the raters with ten reflections, each of which the raters were asked to rate on their own over a 20 minute interval. Following this, the trainers tallied up the responses, and any reflections that had more than a single disagreement were discussed. Discussion is imperative for both the group and individual exercises so that all the raters are calibrated and rate using the same frame of reference. The total training session can last between one and two hours.

Materials

PowerPoint presentations and handouts are available online at http://www.epsilen.com/ddalal under the “Showcase” tab.