

## Using internship results as assessment of accounting outcomes

Henry Elrod  
University of the Incarnate Word

Theresa Tiggeman  
University of the Incarnate Word

### ABSTRACT

Learning outcome goals, established by faculty in business and accounting, include some level of mastery of technology, ability to demonstrate critical thinking, and competence in the fundamental disciplines of business. Syllabi contain outcomes related to the fundamental student learning outcome goals and explicit measures of outcome performance such as tests, homework, research papers, presentations, and student achievement evidenced by certification in specific areas of the curriculum. Validity for measures of competency in the fundamental disciplines of business, in critical thinking, and in the use of business technology requires confirmation by sources independent of direct measures, such as grades, or other faculty assessments. Constant comparative thematic analysis of internship performance reports, and self-assessment reports, together with descriptive statistics and analysis of metrics developed from performance report and self-assessment data were used to build a rubric for future comparative analysis. Performance reports were responses to questions from the employers, and self-assessments were reflection papers prepared by the students. Data from the performance reports and reflection papers were analyzed by coding and counting qualitative descriptors of performance and identifying thematic patterns that seemed to relate to positive or negative assessment. Because the internship participants represented about 75% of the students in the accounting program studied, and the assessments were prepared by the employers and the students, not faculty, the baseline data and assessment rubric constituted an independent third party assessment of accounting student competencies in critical thinking, in the use of technology, and in mastery of the fundamental disciplines of business.

Keywords: outcomes, assessment, internship, accounting, thematic

## INTRODUCTION

If university accounting programs are to prepare students to meet or exceed employer expectations, especially with regard to establishing and evaluating major student learning outcome goals, program faculty and administrators must have information about overall employer attitudes toward program, the goals and the student's levels of achievement (Verney, Holoviak, & Winter, 2009). Student learning outcome goals, normally established by faculty in business and accounting (AACSB, 2011), for the university where this study was conducted include some level of mastery of technology, ability to demonstrate critical thinking, and competence in the fundamental disciplines of business, including accounting. Assessment of student learning outcome goal achievement can be accomplished with both direct and indirect methods (Lusher, 2010). In accreditation learning outcome goal assessment guidelines, there are calls for both direct and indirect assessment methods. For instance, the handbook for development of learning outcome assessment plans at the University of Central Florida delineates an inventory of both direct and indirect assessment methods, and includes discussions of the advantages and disadvantages of each (Selim, Pet-Armacost, Albert, & Krist, 2008). Acceptable indirect assessment methods include survey reporting by employers and student satisfaction surveys. Such surveys can include evaluation of perceived learning outcomes by employers as well as students. Finally, a general constraint applicable to reliable and relevant information intended to be useful to decision makers and other users, is that the information gathering process should cost-effective and provide benefits greater than its cost (Financial Accounting Standards Board, 1980; Nikolai, Bazley, & Jones, 2010).

Validity or credibility and transferability for measures of competency in the fundamental disciplines of business, in critical thinking, and in the use of business technology requires confirmation by sources independent of in-class measures, such as grades, or other faculty assessments, so that results are dependable and reproducible and able to be corroborated (Trochim, 2001). For internships, among other methods, outcomes may be measured by surveying the internship providers or employers, by requiring time and activity records such as time sheets or work diaries, by surveying the interns, and by grading and content analysis of intern reflection papers. To the extent (a) the formative assessment perceptions of the employers/providers, and (b) the reported experiences of the interns can be generalized to the program rather than providing only direct assessment of the individual interns, these course level direct assessments may be used as indirect program level assessments of the student learning outcomes addressed.

Program level assessment may be accomplished by using the individual direct assessment of performance reports from employers/providers, and reflection papers from the interns, as data that, when analyzed, provide indirect assessment evidence for specific program student learning outcome goals. State requirements for accounting internships that may be counted in candidates' accounting courses for purposes of qualification to sit for the licensing examination, are focused on recording evidence that the knowledge gained from an internship is equal or greater than the knowledge gained from a 3-hour course in a traditional accounting classroom setting (TSBPA, 2011), but without regard for whether the knowledge gained, in either situation, contributes toward achievement of the program-level student learning outcome goals. Verney, Holoviak, and Winter (2009), in a similar study focused on the efficacy of internship performance assessment tools, as opposed to a focus on using direct assessment measurements of individual performance as indirect evidence of achievement of student learning outcome goals, asserted that the indirect

evidentiary use of the direct evidence from internship assessment was possible and beneficial. In a study designed to show that internship evaluation provides feedback as to student performance, they also showed such information was useful to program review and revision efforts. They presented evidence that improvement in the internship assessment tools can provide greater variability in the results of the assessment process, and hence greater distinction across the spectrum of interns' performances.

Indirect evidence from intern performance assessment can include both student surveys (Hill, Perry, & Stein, 1998) and student reflection papers, both forms of participant generated assessment. Questions about reliability or validity notwithstanding, the usefulness of self-reported assessments, student surveys are thought to provide useful measurement of student attitudes and student satisfaction, but may not address the level or achievement of program student learning outcome goals. Together with the supervisors' evaluations of internship performance, self-assessment reflection papers can provide useful insights. Beard (1993) reported that in empirical studies of internship benefits, focused on measuring the educational benefits of internships, intern-reported reflections and insights have been under-used. Thematic analysis of the performance professionals in a nursing practice, through self-reported reflection papers can be rigorous, and can present valid, usable data about the professionals' performances (Fereday & Muir-Cochrane, 2006). Application of such approaches and techniques, as described by Aronson (1994), to accounting internships should not present problems.

The problem statement was, as demanded by constituents/stakeholders and accreditation assessment standards, faculty need cost-effective reliable and valid sources of information about the success of programmatic student learning outcomes. The research question was can course-level direct assessment documents, in particular, interns' self-assessment reflection papers, for individual interns in an existing accounting internship curriculum provide cost-effective, reliable and valid of information about the achievement of program level student learning outcomes?

## **STUDY DESIGN**

The internship reporting and assessment process provided direct, performance-based course-related assessments of the performance of each individual intern. For this study, internship evaluation reports, with names and identifying material (such as student ID numbers) redacted from the documents, were used as secondary data to provide indirect assessment evidence about achievement of programmatic student learning outcome goals. Use of the same data for two different purposes makes both processes more cost-effective, answering the demand of accreditation guidelines and of the standards for quality information (Financial Accounting Standards Board, 1980; Nikolai, et al., 2010). Although the research question and problem statement were well settled, as is often the case with qualitative research (Stake, 2010), the methodology for this study developed as the data were examined. The phenomenological inquiry method (Lichtman, 2010), in this study by examination of the self-reported common experience of the accounting internship through reflection papers, provided a research approach that lead to a better understanding and interpretation of the data .

The particular student learning outcome goals of interest in this study were: students will (a) demonstrate knowledge of the fundamental disciplines of business, (b) proficiently use technology to support analysis and decision making skills, (c) demonstrate the ability to think critically and communicate complex ideas, and (d) demonstrate an awareness of domestic and global social issues and the role of business in solving them. Of these the outcome goal d, related

to global social issues and the role of business in their solution, was eliminated from the study by inspection. There were no questions related to this student learning outcome goal in the supervisor assessment form, or the student site assessment form, and none of the student interns addressed such issues in their reflection papers. This does not mean the student learning outcome goal was not achieved. It simply means the instruments and papers used for data in this study did not address the issue. In a similar fashion, the student learning outcome goal related to mastery of the fundamental disciplines of business was also eliminated from the study, for the same reasons cited with regard to the elimination of the global social issues outcome. The forms and papers constituting the evidence for this study simply did not address the issue clearly enough to allow competent analysis.

The student learning outcome goal related to the effective use of technology was assessed by summing the supervisor's evaluation form ratings for question 6, relating to the preparation of quality work within an appropriate time period, and question 8 in the supervisor's evaluation form, relating to the intern's ability to communicate orally and in writing. These questions implicitly include the ability to leverage the technology used in CPA firms to produce quality work by making effective use of industry standard word processing and spreadsheet software and the related supporting computer technology.

In a manner similar to that used for the technology goal, the supervisor's evaluation form score was used to benchmark the interns critical thinking score as measured by summing the numerical ratings given by the supervisor for questions 9 and 12 in the supervisor's evaluation form.

The direct assessment of internship performance documents were the student interns' reflection papers from 53 internships, the related internship supervisors' evaluations, the student internship site evaluation form, and the internship control sheets. Completed at the end of the internship experience, the reflection papers were typed double spaced and generally ran in length from one and a half to three pages, or roughly 250 to 700 words. The supervisor's evaluation sheet contained 12 specific questions and one summary question, each with a Likert scale for which 1 was the maximum score and 5 was the minimum score. The student's site assessment contained seven questions, plus a summary question, with a similar Likert scale. The control sheets contained information about interns' academic classifications, grade-point averages, hours worked, pay rates, and the like.

Comparative thematic analysis of internship self-assessment reports, together with descriptive statistics and analysis of metrics developed from performance reports, site assessments and self-assessment reflection papers were used to build a rubric for future comparative analysis, and to speak to, if not fully answer the research question posed for the study. Performance reports were responses to questions from the employers, and self-assessments were reflection papers prepared by the students. Data from the performance reports and reflection papers were analyzed after coding and counting qualitative descriptors of performance and identifying thematic patterns that seemed to relate to positive or negative assessment.

Descriptive statistics were developed by the application of *t*-tests intended to determine whether intuitive notions about relationships between various scores on the supervisor ratings and descriptive items in the data and from the reflection paper analyses were statistically significant. For instance, one might presume that a very well paid intern would be happier with their internship than a poorly paid or unpaid intern. There were up to seven relationships hypothesized across the demographic and descriptive statistics of the intern data, generally

against achievement assessment as determined by the supervisors' evaluation scores. These descriptive relationship hypotheses are set out in Table 1.

As recommended by Richards (2009), to enhance reliability or validity, two researchers independently examined the performance reports and reflection papers to code the qualitative data, and then the results of the coding process were compared. Differences were resolved by reexamining the source documents, and either revising one or both of the items, or noting the reason for the inconsistency and making slight adjustments to the definitions of the categories. Besides making the data easier to manipulate and interpret, the coding process was intended to ameliorate essential imprecision introduced by the diversity of preparers of the reports

## FINDINGS

### Descriptive Statistics

Documents and scores from 53 internships from the years 2006 through 2011 were included in the study, with 11 from 2011, 12 from 2010, 13 from 2009, 14 from 2008, and the balance scattered across 2006 and 2007. The average grade point average of the students was 3.55 (4.0 scale). Thirty-five, or 66% of the interns were graduate students and the balance were undergraduates. A tiny minority, perhaps three or four, had done internships as undergraduates and as graduate students, and were therefore in the data twice. The minimum hours required by the state for an internship to be accepted as part of the students' accounting education is 140 hours (TSBPA, 2011). The average hours worked by the interns in the study was 297. Across the years 15 internships were in the fall term, 5 were in the summer term, and 33 were in the spring term. The average pay was \$13.61 per hour, including 8 that were unpaid. The range of pay (for those paid) was from the low of \$2.33 per hour to a high of \$33.00 per hour, with a standard deviation of \$8.26.

### *t*-tests

The descriptive statistics related to the *t*-tests of possible relationships are set out in Table 2. These statistics include the sample sizes, sample means, and standard deviations for each of the null hypotheses tested. The null hypotheses for the tests are in Table 1. The results of the tests, including the *t*-values, degrees of freedom, significance level for the tests, and conclusions, are shown in Table 3. The SR numbers are arbitrary, consecutive numbers. As shown in Table 3, the *t*-tests for SR 1 through SR 3, SR 5 through SR 9, and SR 10 and SR 11, did not show a statistically significant difference in the means compared, at the 1% level of significance. This meant that the means of internship supervisor's evaluations of their interns were not different when groups were compared and the groups were classified based on grade point averages, intern pay levels, graduate or undergraduate classification, quality of reflection papers, hours worked, use of positive words and phrases in the reflection papers, whether the interns worked for CPA firms, or by fall or spring term.

Four of the null hypotheses constructed to describe relationships in the data were rejected at the 1% level of significance. *t*-statistics and degrees of freedom are set out in Table 3. SR 4 and SR 12 were related, in that SR 4 hypothesized there was no statistical difference in mean supervisor rating for interns scoring the internship site at 1.0, versus mean supervisor rating for interns scoring the internship site at other than 1.0, whereas SR 12 tested a complimentary null

hypothesis that there was no difference in the mean site rating (from the interns' scores on the student internship site evaluation forms), with the interns classified based on their supervisor rating of 1.0 or other than 1.0. This means that when an internship was not going well, both the intern and the supervisor were aware of it, or became aware of it. This simple test does not indicate causality.

The remaining two of the descriptive null hypothesis tests conducted, SR 8 and SR 9 both produced *t*-values and degrees of freedom that indicated the null hypotheses of equality of the means compared should be rejected, at the 1% level of significance. These values are set out in Table 3. The null hypothesis for SR 8 stated there was no statistical difference in the mean scores on the supervisor's evaluation form for questions related to the intern's ability to effectively utilize technology, where the interns were classified by their overall supervisor evaluation scores. In other words, a low grade from the supervisor on the technology related questions was statistically related to low overall ratings from the supervisor. SR 9 hypothesized no difference in mean scores on the supervisor's evaluation form for questions related to the interns' critical thinking and problem solving skills, when the interns were classified their overall supervisor's ratings. This means that low scores on the critical thinking related questions in the supervisor's evaluation form were statistically related to low overall ratings from the supervisor. Ergo, these two statistical tests indicated that the direct evidence of the supervisor's evaluation forms, intended as summative evaluation of individual intern's performance, can indeed provide indirect programmatic level evidence of achievement or non-achievement of the two student learning outcome goals related to the use of technology and the demonstration of critical thinking.

### **Reflection Papers.**

Direct thematic analysis of the content of the interns' reflection papers was made by categorizing the topics they address in the paper, and counting the frequency of those comments. Student comments related to domestic and international social problems and the role of business in solving them appeared in 5.66% of the student reflection papers. This confirmed the observation made from the other analyses, that the internship and the assessment materials for the internships, do not adequately address this student learning outcome goal. The student learning outcome goal related to student familiarity/mastery of the fundamental disciplines of business was addressed in 92.5% of the reflection papers, indicating that, no matter the level of learning, students were aware of having learned new skills. In 84.9% of the papers, students wrote about learning and using new technology. The student reflection papers content analysis showed that 55.6% of the writers included discussions of connecting their internship experiences to their classroom knowledge or included statements that students thought their critical thinking abilities had been enhanced by the internship experience. This perception by the students confirms the critical thinking scores in the internship supervisors' evaluation forms, in which supervisors rated 61% of the students at the maximum score for the two critical thinking questions.

### **RECOMMENDATIONS**

This research was essentially exploratory. The problem addressed was the demonstrated need, driven by both stakeholder/constituents and accreditation guidelines, for independent evidence of accomplishment of learning outcomes to supplement the direct evidence of

curriculum based and course derived performance related measures (Lusher, 2010; Selim, et al., 2008). The research question was are individual reports of internship performance, from the employers, and self-assessment reflection papers from the student interns, useful as supplemental measures of achievement of student learning outcome goals of the program? The essential epistemological distinction described by Trochim (2001) and by Stake (2010), between qualitative and quantitative research, about the exploration of personal constructs and personally constructed realities (or as Anderson (1990) described, socially-constructed realities ) versus attempts to measure parameters of a single discernable reality external to the perception of the researcher, made the work both interesting and difficult, as the interpretation of the data meant using the measurements of over one hundred different reporters, from the students in their reflection papers to the employer supervisors in their assessments of their intern's performance.

### **Reflection Papers.**

One specific recommendation is drawn from the process of grading or coding the content of the interns' reflection papers. Very little direction was given to the interns as to the purpose, content, or style of these papers, resulting in a wide variety. At inception, the interns were told that at the end of the internship the school would collect evaluations from the supervisors and a reflection paper from the interns. Initially, it was thought that the language of the essays could be related, perhaps statistically, to the supervisor ratings. They were also told that the internship grade would be either satisfactory or unsatisfactory, as opposed to A, B, C type letter grades. It was tentatively posited that essays filled with happy thoughts, words, and phrases, would come from interns receiving the highest of ratings from their supervisors. By corollary, it was thought interns who did not receive high performance ratings would prepare reflection papers with more words of unhappiness, discontent, and frustration than those of their highly rated colleagues. The reality was that the students saw the reflection papers as summative documents, rather than formative documents. Consequently, believing their grade depended on having a happy and successful internship experience, virtually no students reported discernable dissatisfaction with the conduct or content of their internships. The formative nature of the direct assessment reflection papers might be enhanced by requiring two papers, instead of the single paper at the end of the process. First, students could be asked to prepare before the beginning of the experience, a short essay describing their expectations. What do they expect to do? What new skills do they expect to learn? How do they expect to be treated by co-workers, and by supervisors? Then, at the end of the internship, they could be asked for a short summative paper that simply recapped the experience, noted what was learned, and make specific recommendations that might be passed along to future interns. Specifically drawing the interns' attention to formative aspects of the experience at the beginning, and then asking them to write about their experiences from the perspective of self-examination and internally guided improvement, should make the internship experience more valuable to the interns.

As suggested by Weimer (2002) summative and formative evaluation efforts should be separate. If they are to be useful to the students in their processes of integrating classroom and real life experiences and knowledge, formative assessments, especially self-reflection assessment, should not be perceived by the interns as summative. If the reflection papers are intended to provide the students with information about how they are doing in the real-world context, the process should be modified to ask for the reflection papers earlier in the internships, instead of after the working portions of the interns' experience are concluded.

Internship supervisors' evaluation forms.

### **Internship Supervisor's Evaluation Form**

As noted in the description of the study, two of the student learning outcome goals (familiarity/mastery of the fundamental disciplines of business and awareness of international and domestic social problems and the role of business in solving them) were eliminated from the study because the extant evaluation forms simply did not address the issues. This calls for a recommendation to revise and improve the forms used to increase the scope to include questions addressing both issues. This should be done in particular with the supervisors' internship evaluation form, but also should be addressed in the revision of the instruction given to the interns for the preparation of their reflection papers.

### **REFERENCES**

- AACSB. (2011, January 31, 2011). Eligibility procedures and accreditation standards for business accreditation Retrieved August 5, 2011, from <http://www.aacsb.edu/accreditation/standards-2011-revised-jan2011-final.pdf>
- Anderson, w. T. (1990). *Reality Isn't What It Used To Be*. San Francisco: Harper San Francisco Division of Harper Collins Publishers.
- Aronson, J. (1994). A pragmatic view of thematic analysis. *The Qualitative Report*, 2(1). Retrieved from <http://www.nova.edu/ssss/QR/BackIssues/QR2-1/aronson.html>
- Beard, V. K. (1993). *Reflections of public accounting interns: A qualitative analysis of experiential learning*. Ph.D. Dissertation, University of North Dakota, Grand Forks, ND. UMI 9408441 database.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 2-11. Retrieved from <http://ejournals.library.ualberta.ca/index.php/IJQM/index>
- Financial Accounting Standards Board. (1980). Statement of Financial Accounting Concepts No. 2. Norwalk, CT: Financial Accounting Standards Board.
- Hill, N. T., Perry, S. E., & Stein, D. M. (1998). Using accounting student surveys in an outcomes assessment program. *Issues in Accounting Education*, 13(1), 65-78. Retrieved from ABI/INFORM Complete database
- Lichtman, M. (2010). *Qualitative Research in Education*. Los Angeles: SagePublications, Inc.
- Lusher, A. L. (2010). Assessment practices in undergraduate accounting programs. *Journal of Case Studies in Accreditation and Assessment*, 1. Retrieved from <http://www.aabri.com/manuscript/10550.pdf>
- Nikolai, L. A., Bazley, J. D., & Jones, J. P. (2010). *Intermediate Accounting, 11th Edition*. Mason, OH: South-Western Cengage Learning.
- Richards, L. (2009). *Handling Qualitative Data, 2nd Edition*. Los Angeles: Sage Publications Inc.
- Selim, B. R., Pet-Armacost, J., Albert, A., & Krist, P. S. (2008). UCF Academic Program Assessment Handbook: University of Central Florida.
- Stake, R. E. (2010). *Qualitative Research: Studying How Things Work*. New York: The Guilford Press.

- Trochim, W. M. K. (2001). *Research Methods Knowledge Base, 2nd Edition*. Cincinnati, OH: Atomic Dog Publishing.
- TSBPA. (2011). Texas State Board of Public Accountancy Retrieved August 8, 2011, from <http://www.tsbpa.state.tx.us>
- Verney, T. P., Holoviak, S. J., & Winter, A. S. (2009). Enhancing the reliability of internship evaluations. *The Journal of Applied Business and Economics*, 9(1), 22-33. Retrieved from ProQuest ABI/INFORM Complete database
- Weimer, M. (2002). *Learner-Centered Teaching*. San Francisco: Jossey-Bass.



Table 1  
*Relationship hypotheses*

Hypothesis No.	Hypothesis
SR 1	There was no difference in mean GPA for interns receiving 1.0 supervisor ratings, versus mean GPA for interns receiving greater than 1.0 supervisor ratings.
SR 2	There was no difference in mean supervisor rating for interns paid above the mean pay, versus mean supervisor rating for interns paid at or below the mean pay.
SR 3	There was no difference in mean supervisor ratings for interns registered under the undergraduate course number, versus the graduate course number.
SR 4	There was no difference in mean supervisor rating for interns scoring the internship site at 1.0, versus mean supervisor rating for interns scoring the internship site at other than 1.0.
SR 5	There was no difference in mean supervisor rating for interns whose reflection papers were assessed as excellent, versus mean supervisor rating for interns whose reflection papers were assessed as other than excellent.
SR 6	There was no difference in mean supervisor rating for interns who worked more than the mean hours worked by all interns, versus mean supervisor rating for interns who worked the mean hours or fewer than the mean hours worked by all interns.
SR 7	There was no difference in mean supervisor rating for interns who expressed satisfaction in their internships by using more than the mean number of positive words or phrases in their reflection papers, versus students who express less satisfaction in their internships by using the mean or fewer than the mean number of positive words or phrases in their reflection papers
SR 8	There was no difference in mean technology score for interns rated 1.0 by their supervisors, versus mean technology score for interns rated other than 1.0 by their supervisors.
SR 9	There was no difference in mean critical thinking score for interns rated 1.0 by their supervisors, versus mean critical thinking score for interns rated other than 1.0 by their supervisors.
SR 10	There was no difference in mean supervisor rating for interns whose internships were with a CPA firms, versus the mean supervisor rating for interns whose internships were not with CPA firms.
SR 11	There was no difference in mean supervisor rating for interns in the fall term, versus the mean supervisor rating for interns whose internships were in the spring term.
SR 12	There was no difference in mean student internship site rating for interns rated 1.0 by their supervisors, versus student internship site rating for interns rated other than 1.0 by their supervisors.

Table 2  
*Descriptive statistics for t-tests of relationship hypotheses*

SR No.	$n_1$	$n_2$	$\mu_1$	$\mu_2$	$\sigma_1$	$\sigma_2$
SR 1	14	20	3.63	3.55	.368	.374
SR 2	23	25	1.54	1.25	.709	.306
SR 3	38	17	1.40	1.26	.499	.621
SR 4	37	15	1.17	1.82	.247	.760
SR 5	38	14	1.37	1.33	.550	.531
SR 6	23	28	1.22	1.48	.279	.672
SR 7	19	33	1.49	1.28	.670	.443
SR 8	24	28	2.00	3.61	.000	1.31
SR 9	24	28	1.00	1.86	.000	.803
SR 10	22	30	1.31	1.39	.419	.618
SR 11	14	33	1.35	1.36	.691	.505
SR 12	24	28	1.02	1.26	.060	.380

Table 3  
*t-test results for relationship hypotheses*

SR No.	$t$	df	Significance level	Conclusion
SR 1	-0.15	32	1%	Do not reject hypothesis.
SR 2	1.88	29	1%	Do not reject hypothesis.
SR 3	0.87	50	1%	Do not reject hypothesis.
SR 4	-3.27	15	1%	Reject null hypothesis.
SR 5	0.22	24	1%	Do not reject hypothesis.
SR 6	-1.85	37	1%	Do not reject hypothesis.
SR 7	1.23	27	1%	Do not reject hypothesis.
SR 8	-5.97	50	1%	Reject null hypothesis.
SR 9	-5.22	50	1%	Reject null hypothesis.
SR 10	0.62	49	1%	Do not reject hypothesis.
SR 11	-0.05	45	1%	Do not reject hypothesis.
SR 12	-3.24	28	1%	Reject null hypothesis.