

# Improving Student Learning Through the Use of Multi-Source Assessment and Feedback<sup>\*</sup>

Jack McGourty and Peter Dominick, Columbia University  
Mary Besterfield-Sacre, Larry Shuman and Harvey Wolfe  
University of Pittsburgh

**Abstract** – *This paper examines the use of multisource assessment and feedback processes in the classroom and the potential impact on student learning. Grounded in control and goal setting theories, this assessment process provides a means for students to take a proactive role in their learning. Research and practice issues are addressed.*

## Introduction

Recent years have seen academic institutions increasingly incorporate broad competency frameworks and outcome assessment processes in the name of educational innovation and reform. While there are many reasons for these educational interventions, in engineering institutions, one of the key drivers is that newly graduated engineers will need competencies beyond the traditional knowledge of science and basic engineering principles. As professionals, they will require skills to help them function in multidisciplinary teams, work with complex globally-based systems of products and services, and strive towards continuous self-learning. In attempting to utilize these frameworks and assessment processes, many colleges and universities have focused on several challenges. These challenges include clearly defining competencies, strategically linking them to course content and accurately measuring outcomes in relation to them. One issue that has yet to be proven is how these educational innovations effect student learning.

Multi-source assessment is a formal process that provides critical information from several sources, such as peers, self and instructors, on student competencies and specific behaviors and skills, affording the student a better understanding of personal strengths and areas in need of development. The typical process is comprised of gathering evaluative information on a target student from two or more rating sources. The target student also provides self-ratings that are subsequently compared with those from the other sources. Once all ratings are completed, the student receives feedback on the behaviors, skills, and performance being assessed. The student can then interpret the results and make decisions on actions that should be taken based on the information provided. As discussed throughout this paper, variations in the design and implementation of multi-source assessment processes can potentially impact student learning.

The authors' current research shows that assessment processes not only provide valuable data on learning outcomes, but also have an impact on learning itself. For

example, introducing a formal assessment process helps to reinforce the learning objectives established for a specific course. When students are actively involved in their own assessment, they are forced to think about their learning in profound ways. This paper examines the use of multi-source feedback systems and how these processes support the development of specific student learning outcomes. Prior to examining the impact of feedback on increased learning, two theories on feedback are reviewed.

## Theoretical Considerations

Both Control Theory [1] and Goal Setting Theory [2] provide a useful framework for explaining how feedback programs may effect the behavior and skill development of participants. First, both theories view behavior as goal directed. Second, both theories suggest that in order to achieve goals people use feedback to evaluate their performance relative to their goals [3]. The two theories share several other key activities, namely: self-monitoring of behavior; followed by self-evaluation of behavior by comparison to a standard; and then self-reaction in the form of satisfaction or dissatisfaction, which in turn leads to further adjustment of behavior and or the modification of goals [4]. In essence, the theories describe self-regulatory loops in which feedback becomes the basis for evaluation and subsequent adjustment.

From the perspective of Control theory receipt of feedback is critical. Control theory suggests that specific feedback is the basis for identifying goal-feedback discrepancies, which subsequently direct one's attention towards change and improvement. According to Carver & Scheier [5] two things must happen in order for people to change their behavior. First, they must focus attention on their behavior so that they can obtain input for comparisons purposes. Second, they must self-regulate their behavior based upon whether or not they perceive a discrepancy between their observed behavior and their standard/goal.

A multi-source feedback program provides the opportunity for both prerequisites to occur. Attention is directed toward one's own behavior at both the time a person completes the multi-source assessment and feedback instrument and at the time one receives feedback. The receipt of feedback also satisfies the second prerequisite by providing the opportunity for individuals to identify discrepancies between how they perceive their behavior and how constituent others perceive it.

---

<sup>\*</sup> This paper supported in part by National Science Foundation grants EEC-9872498, *Engineering Education: Assessment Methodologies and Curricula Innovations* and EEC-9109794 and EEC-9727413, *Gateway Engineering Education Coalition*.

In Goal Setting Theory, similar outcomes are predicted; however, from this perspective, people are not motivated by the need to reduce the discrepancy but rather by the desire to achieve the goal [6]. In goal setting language individuals either work to achieve the goal, change the goal, reject the feedback or abandon commitment to the goal. However, they are most likely to continue working towards goal achievement when their initial performance is below the goal standard [7]. As is the case with Control Theory, receipt of feedback becomes valuable because it is the means by which people know whether or not their performance is above or below the standard.

### Impact on Student Learning

While it is early in the research, there is growing evidence that feedback processes have a positive impact on student learning and attitudes. In several engineering schools, students have been exposed to multi-source feedback processes as part of their design courses. Working in teams provides the opportunity to receive feedback from peers, teaching assistants, and faculty. Results presented in an earlier paper by McGourty [8] consistently demonstrate that students improve based on perceptions of peers and faculty. In another example, Atwater, Roush & Fischthal [9] found that 'follower' ratings of student leadership skills improved after feedback was given to leaders and that leaders receiving "negative" feedback (defined as self-ratings that were considerably higher than follower ratings) improved the most. Subjects in this study were 978 leaders (juniors) at the U.S. Naval Academy and 1,232 followers (freshman). The Leaders received confidential feedback from the freshman followers they had in their squads. The leaders' feedback was aggregated information based upon two rounds of followers' ratings taken six weeks apart. The dependent measure was a post feedback rating provided some eighteen weeks after the second round of feedback. These post feedback ratings were actually provided by a different group of followers than those who had provided the original feedback. In addition to finding improvements in performance based upon follower ratings, they also found that the leaders' self-ratings became more consistent with follower ratings after receiving feedback.

Introducing a systematic program of assessment and feedback helps to reinforce key learning objectives. First, behaviorally specific information contained in such assessment mechanisms helps to define and make salient to the student what is required in order to perform effectively. Second, the fact that the information is presented as part of a formal feedback system sends a strong message to students that performance should be improved. Recent research on the use of multi-source feedback systems suggests that students are likely to demonstrate changes in behavior and skill acquisition simply by completing the feedback instrument. In a recent laboratory study, Dominick, Reilly and McGourty [10] examined the behavioral impact of peer feedback relative to exposure to the feedback instrument.

Subjects were randomly assigned to teams with four or five members, and teams were randomly assigned to one of three experimental conditions: feedback, exposure, or control. Each team performed two group decision making tasks. In the feedback condition, team members rated themselves and each other using a 24-item behavioral observation scale after completing the first task. The scale was based upon a four dimensional model of team behavior [11]. Prior to performing the second task (one week later), they received individualized feedback reports that summarized how they rated themselves and how their peers had rated them. In the exposure condition, team members completed the same 24-item scale after completing the first task but did not receive individualized feedback until after completion of the second task. Team members in the control condition did not complete the instrument or receive a feedback report. The second task was videotaped and rated by experts blind to experimental condition. Results showed significantly higher means on all four dimensions for participants in the feedback and exposure conditions compared to the control group; there were no differences between subjects in the feedback and exposure conditions.

These results provide a more definitive picture of feedback's role relative to exposure. Rather than relying on participants' perceptions of behavioral change, this study measured behavioral differences using external raters blind to condition, thereby showing that it is actual behavior that is different (not merely perceptions of program participants). In addition, unlike the previous studies, this study included a sample not exposed to the feedback instrument and as a result more clearly identifies exposure as the critical mechanism in creating behavior change.

Within the context of a peer feedback program there are several reasons to believe that behavioral improvements should occur simply through exposure to and completion of the feedback instrument. From a cognitive perspective, information about a behavioral construct (e.g. teamwork) can impact a person's schema for that construct. That schema then helps to shape how people will, on the one hand, evaluate others' teamwork behavior, and on the other hand can also act as a script that guides their own behavior [12].

Similarly, just introducing a feedback instrument provides people with referencing information that may influence their notions of what constitutes effectiveness. As Van Velsor & Leslie [13] note, a feedback instrument consisting of behavioral items is, by nature, provides people with a learning opportunity. By reviewing the items, individuals are exposed to examples of effective behavior and are given a framework for understanding/self-evaluating their behavior in a particular context.

Locke and Latham [14] have argued that the just implementing a formal feedback system into a work group may sometimes cause spontaneous goal setting. By introducing the system a strong message is sent that performance should be improved in those areas being

measured. For example, understanding that effective decision-making includes “anticipating problems and developing contingency plans,” may increase individuals’ efforts to do so in the future. The process of completing self and peer-ratings provides individuals with an opportunity to reflect upon their own behavior and to establish normative standards and personal improvement objectives. Consistent with Control Theory, individuals can then use these standards and objectives as a basis for self regulation. Specifically, individuals who perceive a discrepancy between their behavior and the standard behavior presented, might be most motivated to change their behavior [15, 16, 17].

Knowledge that behavior is being measured and that feedback will occur at some later time may be equally as important as exposure. For example, research demonstrates that repeated administration seems to result in sustaining the initial change over a long period of time [18]. Studies on goal setting suggest that people will revert to lower standards once feedback is removed from a goal setting process [19]. Research also suggests that when feedback recipients hold follow-up discussions with feedback providers, they are more likely to improve when compared to people who just receive feedback but do not follow up [20].

### Psychometric Issues

Multisource assessment instruments used in the classroom have proven to be both reliable and valid. Reliability is commonly defined as the measure of random-error variance. Highly reliable assessment instruments will have little variance among raters within a specific source, such as peers or instructors (Inter-rater) and among correlations of various skill areas being assessed (Internal consistency). Inter-rater reliability for a given student sample typically ranges from .37 to .72. Internal consistency estimates range from .78 to .94 [21]. These results on the reliability of the instrument are similar to findings in industrial applications of multi-source assessment [22].

Results from our classroom research have demonstrated reasonable agreement between different rating sources, for example student self-ratings and peers or peer ratings and faculty. For example, Reilly [23] demonstrated that correlations between student self and peer ratings range from .12 to .39. Lower agreement between rating sources may not be a function of random error variance. Other researchers have suggested that low agreement may be due real behavioral or skill differences in the target student as perceived by sources with different perspectives, fellow students versus instructor [24, 25].

An assessment method’s validity is defined as the degree that the instrument in question measures what it is suppose to measure [26]. Evidence of validity is ascertained in one or more of three methods – content, criterion-related, and construct. Content validity is defined as the degree that the assessment method represents the behavioral, skill, or classroom performance domain. In industry, the content

validation process is based on rigid and systematic task analyses that identify exactly what behaviors and skills are required to effectively perform the task in question. These task analyses are normally performed by subject matter experts; individuals who know the task and the knowledge and skills required to perform the task. Once the critical knowledge and skills are identified, they can be used to develop survey items that will be included in the multi-source feedback process. There are few reported examples of formal content validation processes being conducted in the academic setting to support assessment development. For example, McGourty, Sebastian, and Swart [27] describe the process that faculty and administrators went through in one engineering school to identify specific knowledge and skills required for students to successfully complete the requirements of freshman and senior design courses.

Criterion-related and construct validity are much more difficult to demonstrate. One of the key issues is how to define the criterion by which the ratings will be compared. In many courses, the definition of student performance is difficult to define quantitatively. However, design courses and classes that use computer simulated games to support learning have ample opportunities to define and collect quantifiable performance data. For example, one engineering school has been using management simulation games as an integral part of their capstone courses. The simulation yields information about team performance in areas of effective use of research and development resources, company revenues and profitability, and other productivity metrics. One way to gather evidence for criterion related validity is to investigate whether the data from the multi-source assessment process can differentiate the high and low performance of various student teams.

Construct validity focuses on the degree that the assessment data measures the innate characteristics of the specified domain whether it is cognitive, behavioral, or attitudinal.

Evidence of construct validity can be established by investigating the relationships between the ratings derived from the assessment source in question and other measures of the same psychological, behavioral, or cognitive area. A number of studies are underway by the authors to address this very issue. For example, the University of Pittsburgh, Department of Industrial Engineering is conducting a longitudinal triangulation experiment that started in the fall 1999 [28]. As part of a larger research project [29], they are evaluating the information obtained when multiple methods are used on a cohort of engineering students who are being tracked from the beginning of their sophomore year until graduation. Overall, four different methods for measuring outcomes: questionnaires, multi-source feedback, concept maps, and intellectual development are being examined. The purpose of the study is to triangulate two or more different assessment methods for measuring student learning outcomes to determine the extent that these methods yield consistent, corroborative results.

In the context of student learning, one way to examine validity is to compare student ratings with those of trained observers. In one class sample, graduate psychology students were trained to observe specific behaviors in a team-based project context. Thirty students worked in teams of six. A designated graduate student observed each team at work. Correlations between these trained observers and student peer ratings ranged from .48 to .61. Average convergent and discriminant validity were .55 and .34 respectively [<sup>30</sup>].

### **Implications for Classroom Practice**

Based on current research, there are several implications for classroom implementation. However, before integrating such multisource assessment processes into the curriculum, several classroom practices should be considered. The first practice to be addressed is the linking of the multisource assessment process to intended student learning outcomes. As research shows, multisource assessment can be a mechanism for communicating key course objectives and learning outcomes to the students. Therefore, it is important that instructors clearly define course objectives and outcomes as part of the course design. These specified objectives and outcomes then must serve as a basis for multisource assessment survey items. In turn these survey items must be constructed to describe observable behaviors and skills for the process to valid. Currently there are investigations on the efficacy of using pre-formulated item databases linked to the ABET EC 2000 program outcomes [<sup>31</sup>]. These databases are designed to help faculty in converting course objectives and outcomes into useful survey items to support classroom assessment activities.

The second critical implementation issue is to determine up front what communication needs to be provided to the students. It is clear from research on feedback processes that certain information should be provided to students prior to administering a multisource survey. For example, studies show that information about rating biases provided to the rater's ahead of time improves instrument reliability and validity [<sup>32</sup>]. Students, who are aware of normal biases that can effect their ratings, are more likely to avoid such rating decision errors. Additionally, students should have a clear understanding of the purpose, benefits, and expected outcomes of the survey process. An important part of this upfront communication is a clear statement as to how results will be used and by whom. For example, students may inflate ratings if they believe that the ratings will impact their grades. This is especially true, if the process of grading is not fully understood. In general, pre process communication can take many forms including full descriptions in course syllabi, classroom handouts or websites, and lectures.

A clear implication from the research is that students need a structured post-feedback process if significant learning is to occur. While exposure to the process may lead to some improvement as discussed earlier, significant learning and change takes place when students focus on the

feedback they receive and the implications to their own development. Typical multisource assessment processes provide some type of feedback report providing students with mean ratings from the various sources on each learning outcome being measured. Feedback reports sometimes compare ratings from various sources and normative data, such as average class results, when available. Additionally, item variance may be provided, allowing students to see the range of ratings and degree of agreement within rating sources. Finally, report data can be presented in ways to help the student focus on strengths and areas in need of improvement. For example, the highest and lowest rated learning outcomes can be identified. Students can begin to investigate ways to develop skills perceived as needing improvement as well as capitalize on perceived strengths.

Students need to understand what to do with results, once provided. Best practice indicates that some structured process should be provided to support students' use of the feedback. One example is to provide students with an assignment in which they develop an action plan on what areas they plan to specifically improve. Students review results, decide on areas in need of improvement, and decide on developmental activities that plan to engage in to support future learning. Again, instructors may want to provide some general information on suggested developmental activities for intended learning outcomes. These suggestions may be as simple as reading a book on the subject or joining a school activity that allows the student to practice the skill in question.

One final issue that needs thorough investigation is whether or not information derived from multisource assessment should be used for grading purposes. There is limited research investigating the potential use of multisource ratings as input into the student's grade. Results from some of the authors' work have shown relatively low, but significantly positive correlations between self and peer assessments and grades given by faculty. Based on student and faculty interviews, the relationship to grades may be due to the tendency to focus on technical competency when grading student performance. Research from industrial applications of multisource feedback and various human resource decisions has clearly demonstrated that ratings inflation is a major problem when raters perceive that results will impact pay or career-related decisions [<sup>33</sup>].

### **Conclusion**

As educators increase their attention to developing and assessing student learning outcomes, multi-source feedback systems will play an integral part in the education process. Literature and current research show that student can play an active role in their own development and assessment. However, these self-assessment skills need to be developed by providing a structured process to facilitate their learning.

## References

- <sup>1</sup> Carver, C.S. and Scheier, M. F. (1981). Attention and Self-regulation: A Control Theory Approach to Human Behavior. New York: Springer-Verlag.
- <sup>2</sup> Locke, E.A. and Latham, G.P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.
- <sup>3</sup> Kluger, A. and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*.
- <sup>4</sup> Kanfer R. (1990). Motivation theory and industrial and organizational psychology. In MD Dunnette and LM Hough (Eds.), *Handbook of Industrial/Organizational Psychology, 2<sup>nd</sup> Edition*, (pp.75-170). Palo Alto, CA: Consulting Psychologists Inc. Press.
- <sup>5</sup> Carver, C. S. and Scheier, M. F. (1981). Ibid.
- <sup>6</sup> Kluger, A. and DeNisi, A. (1996). Ibid
- <sup>7</sup> Locke, E.A. and Latham, G.P. (1990). Ibid
- <sup>8</sup> McGourty, J. (2000). Assessing and enhancing student learning through multi-source feedback. *IEEE Transactions in Education*, In press.
- <sup>9</sup> Atwater L.E., Roush P., and Fischthal A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology*, 48, 35-59.
- <sup>10</sup> Dominick, P.G., Reilly, R.R., McGourty, J. (1997). The effects of peer feedback on team member behavior. *Group and Organization Management*, 22, 508-520.
- <sup>11</sup> McGourty, J., DeMeuse, K.P., and Dominick, P. (1994). Self-managed work teams: A conceptual model for practice and research. Paper presented at the Academy of Management Annual Meeting, Dallas, TX.
- <sup>12</sup> London M, and Smither, J.W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803-839.
- <sup>13</sup> Van Velsor E, Leslie, J.B. (1991). Feedback to managers, Volume II: A review and comparison of sixteen multi-rater feedback instruments. Greensboro, NC: Center for Creative Leadership.
- <sup>14</sup> Locke, E.A. and Latham, G.P. (1990). Ibid.
- <sup>15</sup> Atwater, L.E., Roush, P., and Fischthal, A. (1995). Ibid.
- <sup>16</sup> Smither, J.W., London, M., Vasilopoulos, N.L., Reilly, R.R., Millsap, R.E., and Salvemini, N. (1995). An examination of the effects of an upward feedback program over time. *Personnel Psychology*, 48(1), 1-32.
- <sup>17</sup> Reilly, R.R., Smither, J.W., and Vasilopoulos, N.L. (1996). A longitudinal study of upward feedback. *Personnel Psychology*, 49, 599-612.
- <sup>18</sup> Reilly, R.R., Smither, J.W., and Vasilopoulos, N.L. (1996). Ibid.
- <sup>19</sup> Ludwig, T.D., and Geller, E.S. (1997). Assigned versus participative goal setting and response generalization: Managing injury control among pizza deliverers. *Journal of Applied Psychology*, 82, 253-261.
- <sup>20</sup> Walker, A.G. (1997). Upward feedback: Incremental improvement in manager's performance over five years. Paper presented at the Twelfth Annual Conference of the Society for Industrial and Organizational Society, St. Louis, MO.
- <sup>21</sup> McGourty, J. (2000). Ibid.
- <sup>22</sup> Conway, J.M. and Huffcutt, A.I. (1996). Testing assumptions of 360-degree feedback: A meta-analysis of supervisory, peer, subordinate, and self-ratings. Paper presented at the Eleventh Annual Conference of the Society for Industrial and Organizational Society, San Diego, CA.
- <sup>23</sup> Reilly, R.R. (1996). Team feedback. Paper presented at the Eleventh Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, California.
- <sup>24</sup> Conway, J.M. and Huffcutt, A.I. (1996). Ibid.
- <sup>25</sup> Tornow, W.W. (1993). Perspective or Reality: Is multi-perspective measurement a means or an end? *Human Resources Management*, 32, 221-229.
- <sup>26</sup> Guion, R.M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11, 385-398.
- <sup>27</sup> McGourty, J., Sebastian, C., and Swart, W. (1998). Development Of A Comprehensive Assessment Program In Engineering Education. *Journal of Engineering Education*, Vol. 87, No. 4. 355-361; (1998).
- <sup>28</sup> Besterfield-Sacre, M., Shuman, L. Wolfe, and McGourty, J. Triangulating assessments: Multi-source feedback systems and closed form surveys. Work in progress.
- <sup>29</sup> Besterfield-Sacre, M., Shuman, L. Wolfe, H., Atman, C. McGourty, J., Miller, R., Olds, B., & Rogers, G. (2000). Matching assessment methods and outcomes: Definitions and research questions. *American Society for Engineering Education Conference Proceedings*, June, 2000 (CD ROM).
- <sup>30</sup> Reilly, R.R. (1996). Ibid.
- <sup>31</sup> "Defining the Outcomes: A Framework for EC 2000" *IEEE Transactions on Education* (May 2000, in press)
- <sup>32</sup> Woehr, D. J. and Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.
- <sup>33</sup> Timmreck, C.K., and Bracken, D.W. (1996). Multisource assessment: Reinforcing the preferred "means" to the end. Paper presented at the Eleventh Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, California.