

# Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education

Ludmila Praslova

Received: 7 July 2009 / Accepted: 11 May 2010 /  
Published online: 25 May 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** Assessment of educational effectiveness provides vitally important feedback to Institutions of Higher Education. It also provides important information to external stakeholders, such as prospective students, parents, governmental and local regulatory entities, professional and regional accrediting organizations, and representatives of the workforce. However, selecting appropriate indicators of educational effectiveness of programs and institutions is a difficult task, especially when criteria of effectiveness are not well defined. This article proposes a comprehensive and systematic approach to aligning criteria for educational effectiveness with specific indicators of achievement of these criteria by adapting a popular organizational training evaluation framework, the Kirkpatrick's four level model of training criteria (Kirkpatrick 1959; 1976; 1996), to assessment in Higher Education. The four level model consists of *reaction*, *learning*, *behavior* and *results* criteria. Adaptation of this model to Higher Education helps to clarify the criteria and create plans for assessment of educational outcomes in which specific instruments and indicators are linked to corresponding criteria. This provides a rich context for understanding the role of various indicators in the overall mosaic of assessment. It also provides Institutions of Higher Education rich and multilevel feedback regarding the effectiveness of their effort to serve their multiple stakeholders. The importance of such feedback is contextualized both in the reality of stakeholder pressures and in theoretical understanding of colleges and universities as open systems according to the systems theory (Katz and Kahn 1966). Although the focus of this article is on Higher Education, core principles and ideas will be applicable to different types and levels of educational programs.

**Keywords** Assessment · Evaluation · Program evaluation · Higher Education · Education · Criteria

---

L. Praslova (✉)  
Vanguard University of Southern California, 55 Fair Drive, Costa Mesa, CA 92626, USA  
e-mail: lpraslova@vanguard.edu

## 1 Introduction

Assessment of educational outcomes plays an increasingly important role in Higher Education; it has been a focus of many recent discussions in the literature as well as in everyday practice of colleges and universities. Specifically, accrediting organizations as well as governments place growing importance on student academic learning, such as content learning and intellectual development, as an outcome of educational programs (Allen 2006; Bers 2008; Brittingham et al. 2008; Ewell 2001, 2006). In addition, the accreditors, governments and workforce representatives expect that institutions of Higher Education appropriately prepare students for the labor force through development of relevant skills and competencies (Toutkoushian 2005; Voorhees and Harvey 2005). Achievement of such outcomes needs to be appropriately documented through the process of assessment.

Unfortunately, assessment is also a source of many frustrations for institutions that struggle to make sense out of various requirements, approaches, and necessary pieces of evidence (Allen 2006; Brittingham et al. 2008; Ewell 2001). However, the process of assessment does not need to be frustratingly complicated. This article proposes a straightforward, systematic and practical approach to assessment by adapting a popular framework used for evaluation of training in business organizations, the Kirkpatrick's four level model of training criteria (Kirkpatrick 1959, 1976, 1996), to assessment in Higher Education. Adaptation of this model to Higher Education helps to clarify the criteria and to create plans for assessment of educational outcomes in which specific instruments and indicators can be linked to corresponding criteria. This article will provide theoretical basis for application of the four level model to Higher Education, examples of such application, and a case study of successful use of this model by a private university in California. This university previously struggled with its assessment efforts because, although such efforts were present, they lacked systematic approach, clarity of purpose, and alignment between educational outcomes and methods of assessment. Adaptation of the Kirkpatrick's four level model helped rectify these problems and meaningfully aligned methods of assessment with desired educational outcomes and with the overall mission of the university.

Although the presented approach is rooted mostly in the US American experience, increasing emphasis on assessment in Higher Education is an international phenomenon (Karpenko et al. 2009; Ewell 2001; Voorhees and Harvey 2005). Thus, this approach will likely be of interest to Institutions of Higher Education around the globe.

### 1.1 Definitions and purposes of assessment

*Definitions of assessment* The term assessment is used in various contexts and has somewhat different connotations. For example, it is commonly used to describe the processes used to certify individual students or even to award grades (Ewell 2001). On the other hand, for accreditation purposes, assessment refers to the collection and use of aggregated data about student attainment to examine the degree to which program or institution-level learning goals are being achieved (Ewell 2001). Thus, assessment takes place at multiple levels: the classroom, course, program, general education, and institution (Bers 2008).

As defined by Allen (2006), assessment is an ongoing process designed to monitor and improve student learning (p.1). Ewell (2006) suggests that “Assessment comprises a set of systematic methods for collecting valid and reliable evidence of what students know and can do at various stages in their academic careers . . . governed by formal statements of student learning outcomes” (Ewell 2006, p. 10). However, current emphasis on assessment also implies that the evidence of student learning is used as the crucial evidence of quality of programs in Higher Education. Growing importance of such evidence appears to be a trend around the world.

*Stakeholder emphasis on assessment* Specifics of functions and responsibilities of Higher Education Institutions, such as the relative importance of teaching students, producing research, and participating in community, are influenced by the type of the institution and by local and national contexts (Toutkoushian 2005). Nevertheless, across these various contexts there is a pronounced and growing pressure to present evidence of educational effectiveness to various stakeholders, including students, parents, governmental and local regulatory entities, professional, regional and national accrediting organizations, and representatives of the workforce (Allen 2006; Ewell 2001; Toutkoushian 2005).

Growing stakeholder interest in assessment appears to be a global phenomenon. In the US American context, the Department of Education (DOE) and regional accrediting organizations are taking an increasingly active approach in requiring institutions to provide evidence of student learning (Ewell 2001). Qualification systems in the United Kingdom, Australia, and Denmark aim to specify and measure competencies of students which also serves as a mechanism for evaluating institutional quality (Voorhees and Harvey 2005). The Russian Ministry of Education and Science has also called for assessment of the results of instruction and is setting standards for the competencies that the graduates of colleges and universities must acquire (Karpenko et al. 2009). Overall, it appears that various stakeholders increasingly use the evidence of student learning (rather than inputs into the system, such as library holdings) to assess quality of educational programs (Allen 2006). Evidence of student learning (examinations, test results, performances, etc.) is increasingly viewed as the primary marker of quality of academic programs (Ewell 2001).

*Assessment as vital institutional feedback* According to Allen (2006), an external requirement for continuous assessment is only one of the reasons that underlie the growing importance of assessment. Perhaps an even more important reason is the overall movement of Higher Education toward being learning-focused and emphasizing student outcomes, as opposed to being teaching-focused. Thus, assessment of educational outcomes is not something that should only be done to satisfy external stakeholder requirements. Assessment of student learning is also a way for Institutions of Higher Education to receive feedback regarding the effectiveness of their core educational mission. If student learning truly is a goal and a focus of education, then assessment of student learning provides vital information that allows Institutions of Higher Education to monitor the effectiveness of their programs and success in accomplishment of their core task. The importance of such feedback to institutions cannot be overstated; its role can be theoretically

contextualized by grounding our understanding of the vital function of assessment in the systems theory (Katz and Kahn 1966).

According to the systems theory, institutions of Higher education, like other social organizations, can be understood as open systems connected to their environment in multiple ways, including input, output, and feedback (Katz and Kahn 1966). Information regarding organizational or institutional functioning in relation to the environment, in the form of feedback, is essential to adjustment and making needed changes and thus to proper functioning and, ultimately, to the survival of the system (Katz and Kahn 1966). Interestingly, and independent of this work, Hansen (1994) applied systems theory to understanding of change in education, specifically to change in public school system in USA.

Application of the systems theory to assessment in Higher Education suggests that because the feedback loop is an essential part of healthy system functioning, institutions of Higher Education aiming to remain relevant in a rapidly changing world need to evaluate outcomes of their work and use the results of such evaluation in the process of continuous readjustment of their programs. Although institutions of Higher Education could also receive system feedback in the form of negative media attention, declining student enrollments or governmental sanctions, it is likely that most institutions and individuals would find feedback in the form of deliberate, continuous self-assessment much more preferable. Thus, the importance of continuous assessment can be contextualized both in the reality of high stakeholder expectations of the Institutions of Higher Education, and in theoretical understanding of organizations as outlined by the systems theory.

## 1.2 Institutional struggles with assessment

Despite both external and internal importance of assessment, many colleges and universities still struggle with understanding assessment and using assessment results to improve learning and teaching (Bers 2008), as well as with technical aspects of assessing student outcomes, including clarification of learning criteria and selecting appropriate measures and instruments (Allen 2006; Bers 2008; Brittingham et al. 2008; Ewell 2001). Moreover, even accrediting agencies are not always clear in their expectations with regard to assessment (Ewell 2001). Institutions of Higher education, faced with pressing, yet not always well defined assessment demands, may hastily select and use various indicators and instruments, such as student evaluations of teaching and alumni surveys, or national standardized achievement tests and locally graded student portfolios, without systematic connection between indicators and criteria to be measured, and without proper contextualization of the overall effort in student and institutional interests.

Such institutions often have difficulty engaging in a sustainable, meaningful program of assessment that is perceived to be of value to the institution itself. However, such institutions can benefit from creating assessment programs that are rooted in a clear understanding of educational purposes and are based on well defined criteria linked to specific indicators or examples of evidence of educational effectiveness.

Overall, there appears to be a need for development of a convenient, logical framework that would help institutions of Higher Education to be more systematic, purposeful and proactive in their assessment efforts, with the goal of creating institutional plans of assessment that will allow collection of appropriate information to satisfy both external assessment demands and internal need for feedback.

### 1.3 Overview of proposed approach to assessment

This article proposes a comprehensive and versatile approach to systematically aligning multiple criteria for educational effectiveness and indicators of achievement of these criteria by adapting a popular framework for evaluation of organizational training, the Kirkpatrick's four level model of training evaluation criteria (Kirkpatrick 1959, 1976, 1996), to assessment in Higher Education. This model provides a rich context for understanding of the place and the role of various instruments and indicators as tiles in the overall mosaic of assessment, as specific indicators are mapped onto four levels of criteria—*reaction*, *learning*, *behavior*, and *results* criteria. Application of the four level model also allows institutions to obtain feedback regarding the effectiveness of their educational efforts that is more specific and differentiated, and thus, from the point of view of systems theory, more useful for organizational change and adjustment.

The four level model is a classic framework for assessing training effectiveness in organizational contexts. Although new models have been proposed, the four level model of training evaluation and criteria continues to be the most popular and often cited (Arthur et al. 2003a; Salas and Cannon-Bowers 2001). Alliger et al. (1997) proposed some augmentation to the framework and further refined terminology and criteria, for example, by referring to behavior criteria as transfer criteria, and by specifying affective reactions and utility judgments as subtypes of reaction criteria. Nevertheless, the original model remains widely used and appears to find new applications in additional contexts.

In addition to traditional use in business and organizational contexts, Kirkpatrick's model has been recently applied to understanding of Higher Education as well. Arthur et al. (2003b) applied the four level model to evaluation of teaching effectiveness in Higher Education, but only used the criteria for levels one and two (reaction and learning), which are very similar across all training or educational settings. Applying the other two criteria (behavior and results) to Higher Education requires some adaptation of the model to the specific context and purposes of colleges and universities. The following section will describe the four level model in more detail and discuss how behavior and results criteria can be conceptualized in Higher Education. It will also outline and illustrate how the four level model overall can be used in the context of Higher Education.

## 2 Adaptation of the four level model of training evaluation criteria to assessment in Higher Education

The four levels of evaluation in Kirkpatrick's model are *reaction* criteria, *learning* criteria, *behavior* criteria, and *results* criteria (Kirkpatrick 1959, 1976, 1996).

Reaction and learning criteria are considered *internal*, because they focus on what occurs within the training program. Behavioral and results criteria focus on changes that occur outside (and typically after) the program, and are thus seen as *external* criteria. It is also useful to keep in mind that external criteria are likely to be influenced by factors other than learning, such as larger organizational or economic contexts (Alliger et al. 1997; Arthur et al. 2003a; Kirkpatrick 1959, 1976, 1996; Landy and Conte 2007).

### 2.1 Reaction criteria

Reaction criteria are trainees' perceptions of training (Kirkpatrick 1959, 1976, 1996). Alliger et al. (1997) proposed the distinction between trainee's reports regarding how much they enjoyed the training (affective reactions) and how much they believe they have learned (utility judgments) within the reaction criteria. Reaction level criteria in Higher Education are represented by student evaluations of instruction and by self-report regarding perceived educational gains.

Many researchers have pointed out the lack of relationship between reaction criteria and the other three levels of criteria (learning, behavior and results), and the meta-analytic study by Alliger et al. (1997) found no relationship between affective reactions and other levels, and only a weak relationship between utility judgments and the other levels of criteria. However, despite the fact that many researchers caution against the use of reactions alone for the assessment of learning, reaction level criteria remain the most often assessed (Alliger et al. 1997; Arthur et al. 2003a; Dysvik and Martinsen 2008; Van Buren and Erskine 2002). According to Arthur et al. (2003a), one of the likely reasons for the wide use of reaction level measures is the ease of collection. However, most researchers also agree that learning, behavior and results criteria need to be measured in order to accurately evaluate training outcomes.

### 2.2 Learning criteria

Learning criteria are measures of the learning outcomes, typically assessed by using various forms of knowledge tests, but also by immediate post-training measures of performance and skill demonstration in the training context (Alliger et al. 1997). In the classroom situation, pre and post tests provide the most direct measure of learning, and such measures are typically used in Higher Education settings (Arthur et al. 2003b). However, it is also possible to use writing samples, performances, speeches, and other class-appropriate assessments. Alliger et al. (1997) proposed specifying immediate knowledge, knowledge retention and behavior/skill demonstration measured within training as subtypes of learning criteria, but this idea received relatively limited support.

### 2.3 Behavioral criteria

Behavioral criteria are also referred to as transfer criteria, a terminology change proposed by Alliger et al. (1997). This level includes measures of actual on-the-job performance and can be used to identify the effects of training on work performance.

In organizations, behavioral criteria are typically operationalized as supervisor ratings or objective indicators of performance such as job outputs (Alliger et al. 1997; Arthur et al. 2003a; Landy and Conte 2007). In Higher Education settings, such criteria may include the evidence of student use of knowledge and skills learned in previously taken classes in their following class work, including research projects or creative productions, in application of learning during internship, and in other behaviors outside the context in which the initial learning occurred.

Post-training behavior can be also operationalized as the workplace or civic behavior. Halpern and Hakel (2003) in their discussion of learning in university and beyond stress the importance of transfer in education and the need to teach students in a way that they will be prepared for unpredictable life tests in the future, outside of the classroom contexts, rather than just classroom tests.

Although learning criteria and behavioral criteria conceptually are expected to be related, research has found relatively modest relationship between the two (Alliger et al. 1997; Arthur et al. 2003a). This is typically attributed to the fact that post-training environments may or may not provide opportunities for the learned material or skills to be demonstrated (Arthur et al. 2003a). This potential constraint needs to be considered in design of assessment instruments, and in collection and interpretation of behavioral data.

#### 2.4 Results criteria

Results criteria are both highly desirable and most difficult to evaluate. In organizational settings, they are operationalized by productivity gains, increased customer satisfaction, increased employee morale following management training, or increase in profitability of organizations (Arthur et al. 2003a; Landy and Conte 2007). Results are often difficult to estimate and results criteria are used considerably less frequently than assessments of any other level of Kirkpatrick's model. Alliger et al. (1997) caution that organizational constraints substantially limit opportunities for collecting results data and remind that sponsors of training may have unrealistic expectations with regard to results level outcomes. Organizational, social and economic constraints greatly influence not only data collection, but the very outcomes on the results (and to some extent behavioral) levels of the four level model.

*Results criteria in Higher Education and multiple stakeholders of education* In the context of Higher Education, prior to establishing results level criteria we need to understand who is to benefit from education. According to Toutkoushian (2005), institutions of Higher Education play a vital role in increasing the human capital of individuals as well as society as a whole. Thus, it appears that there are at least two parties that are to profit from education: a) the student, who should develop skills useful for the workplace and life in general, and b) the society, which is interested in college graduates who are competent and responsible contributors to local and global communities.

Benefits to society can be understood in several ways. In the eyes of many stakeholders, colleges and universities first and foremost serve to provide training for students to enter the labor force (Ewell 2001; Toutkoushian 2005). However,

society also benefits from research and service activities and from having a more highly-educated population overall (Toutkoushian 2005). Moreover, although economic benefits to local and national economies through greater productivity of the workforce are extremely important, another, and perhaps even more important way in which colleges and universities contribute to society is through character development and ethical readiness of college graduates for important leadership and civic roles in society (Boyer and Hechinger 1981; Colby et al. 2003; Dalton et al. 2004; Russell 2004; Saxs 2004).

Thus, results criteria in education may include a wide range of outcomes, such as alumni employment and workplace success, graduate school admission, service to underprivileged groups or work to promote peace and justice, literary or artistic work, personal and family stability, and responsible citizenship. Moreover, most of these outcomes benefit both individual and the society. For example, recent work by Biesta (2009) on the purpose or purposes of education largely deals with outcomes which would be classified as results criteria in the four level model, and outlines three important broad functions of education: qualification (providing skills), socialization, and subjectification (preparation for independent thinking and action). All three functions influence both individuals and society, which further illustrates the multiple bases of the importance of good education.

In sum, the four levels of criteria of training evaluation in organizational settings appear to have clear parallels in Higher Education. Table 1 summarizes the four level model in its original application to organizations, outlines possible adaptation to Higher Education settings, and provides some examples of linking specific instruments and indicators to corresponding criteria. In addition, a brief case study presented in the next section will further illustrate how the model can be applied in Higher Education.

**Table 1** Four level model of evaluation criteria applied to training in organizations and to Higher Education

| Criteria              | Training in organizations   | Learning in Higher Education  | Sample instruments and indicators for Higher Education  |
|-----------------------|---|---|---|
| Reaction              | trainee affective reactions and utility judgments   | student affective reactions and utility judgments   | student evaluations of instruction [insert specific instrument]   |
| Learning              | direct measures of learning outcomes, typically knowledge tests or performance tasks                              | direct measures of learning outcomes, knowledge tests, performance tasks or other graded work                   | national or institutional pre-and-post tests [insert specific test]<br>national standardized field test [insert specific test]<br>examples of class-specific student work |
| Behavior/<br>transfer | measures of actual on-the-job performance: supervisor ratings or objective indicators of performance/job outputs. | evidence of student use of knowledge and skills learned early in the program in subsequent work, e.g., research | end-of-program integration papers or projects, internship diaries, documentation of integrative research work,  |



**Table 1** (continued)

| Criteria | Training in organizations   | Learning in Higher Education   | Sample instruments and indicators for Higher Education   |
|----------|---|--|--|
|          |   | projects or creative productions, application of learning during internship, development of a professional resume, and other behaviors outside the context in which the initial learning occurred. | documentation of community involvement projects, and other materials developed outside the immediate class context                   |
| Results  | productivity gains, increased customer satisfaction, employee morale for management training, profit value gained by organization | alumni career success, graduate school admission, service to society, personal stability.  | alumni surveys, employer feedback, samples of scholarly or artistic accomplishments, notices of awards, recognition of service, etc. |

### 3 Clarification of assessment criteria and Indicators through adaptation of Kirkpatrick's four level model to Higher Education: A case study

The usefulness of the adapted four-level model for aligning assessment criteria and indicators can be illustrated by the example of Vanguard University of Southern California, a small private university aiming to have a long-term positive impact on the lives of students and on society. During the 2007–2008 academic year, the university conducted a self-study which included evaluation of academic programs and a close look at assessment of student learning. During this study it became clear that while students were indeed learning and a variety of ways to collect evidence of learning were used, it was difficult to present and interpret the outcomes of learning as evidence of program effectiveness in a concise and systematic manner and to present data in a way that would satisfy both external requirements and internal need for meaningful institutional feedback. This was especially true with regard to evaluation of the university-wide general education program required of all students (core curriculum), although evaluation of departmental programs could also be improved through clarification of criteria for learning outcomes and specific indicators of achievement (De Roulet et al. 2009).

During the 2008–2009 academic year various committees addressed the challenge of improving assessment and evaluation, and a Taskforce for assessment of the core curriculum was assembled with the goal of further evaluating effective and ineffective aspects of existing assessment practices and creating a more functional system for evaluation. The assessment Taskforce thoroughly reviewed the assessment literature and considered multiple examples of approaches to assessment used by other colleges and universities (De Roulet et al. 2009). Although many interesting and useful models were identified, none of the examples fully aligned with Vanguard's educational mission, which reaches beyond the immediate learning and aims to have a positive, long-term

impact both on the student's life and on the workforce, local community, and society in general. While other models fell short of effectively conceptualizing and guiding the assessment of such long-reaching goals, Kirkpatrick's 4-level model could be meaningfully adapted to this end.

The Taskforce mapped specific assessment instruments, such as knowledge tests, samples of student work, and student and alumni surveys onto the model, which provided a rich context for the plan of assessing university educational outcomes. In addition to providing a versatile framework for internal use, this adaptation of the four-level model was positively received by the regional accrediting agency. Currently, the university continues to use the modified Kirkpatrick framework to conceptually guide and inform its effort of continuous assessment and educational improvement.

In conclusion, the use of the adapted Kirkpatrick model in the context of Higher Education can provide colleges and universities both a versatile tool for creating and refining their evaluation and assessment systems, and a way to contextualize both short-term and long-term organizational outcomes, beyond immediate reactions to a specific class (which may or may not correspond to the utility of this class to student future success), and beyond scores on specific standardized tests. Although data for behavior and especially results criteria might be difficult to obtain and such data will rarely be complete, when available, such data is uniquely useful to evaluation and understanding of program outcomes. Moreover, consideration of multiple levels of criteria is a useful reminder of ultimate purposes of instructional and co-curricular efforts.

Although the focus of this article is on colleges and universities and most examples are rooted in student experiences in the US American system of Higher Education, the four level criteria model is likely to be applicable to other types of educational programs and to different national and multinational contexts. If educational institutions are conceptualized as social systems in systems theory, they all need feedback, which in education takes the form of assessment, in order to thrive and to survive. Consideration of reaction criteria, learning criteria, behavior criteria, and results criteria provides feedback that is rich, fine-tuned, multilevel, and considers not only immediate but also long-term outcomes. Such feedback is likely to be most useful to educational institutions as they strive to effectively serve their multiple stakeholders, including students, the workforce, and the overall society.

## References

- Allen, M. J. (2006). *Assessing general education programs (2006)*. Bolton: Anker.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of relations among training criteria. *Personnel Psychology, 50*, 341–358.
- Arthur, W., Jr., Bennett, W. J., Edens, P. S., & Bell, S. T. (2003a). Effectiveness of training in organizations: a meta-analysis of design and evaluation features. *The Journal of Applied Psychology, 88*, 234–245.
- Arthur, W., Jr., Tubre, T. C., Paul, D. S., & Edens, P. S. (2003b). Teaching effectiveness: the relationship between reaction and learning criteria. *Educational Psychology, 23*, 275–285.
- Bers, T. H. (2008). The role of institutional assessment in assessing student learning outcomes. *New Directions for Higher Education, 141*, 31–39. doi:10.1002/he.291.

- Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21, 33–46. doi:10.1007/s11092-008-9064-9.
- Boyer, E. L., & Hechinger, F. M. (1981). *Higher learning in the nation's service*. New York: Carnegie Foundation for the Advancement of Teaching.
- Brittingham, B., O'Brien, P. M., & Alig, J. (2008). Accreditation and institutional research: the traditional role and new dimensions. *New Directions for Higher Education*, 141, 69–76. doi:10.1002/he.294.
- Colby, A., Ehrlich, T., Beaumont, E., & Stephens, J. (2003). *Educating citizens: Preparing America's undergraduates for lives of moral and civic responsibility*. San Francisco: Jossey-Bass.
- Dalton, J. C., Russell, T. R., & Kline, S. (2004). Editors' notes. *New Directions for Institutional Research*, 122, p1–p2.
- De Roulet, D., Praslova, L., Carmody, T., Roberson, M., & Wilson, M. (2009). Assessing core curriculum. Working document, Vanguard University of Southern California.
- Dysvik, A., & Martinsen, Ø. L. (2008). The relationship between trainees' evaluation of teaching and trainee performance among Norwegian executive students. *Educational Psychology*, 28, 747–756. doi:10.1080/01443410802259253.
- Ewell, P. T. (2001). Accreditation and student learning outcomes: A proposed point of departure. *Council for Higher Education Accreditation*, Washington, DC.
- Ewell, P. T. (2006). *Making the grade: How boards can ensure academic quality*. Washington: Association of Governing Boards of Colleges and Universities.
- Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond: teaching for long-term retention and transfer. *Change*, July/August, 2–13.
- Hansen, J. B. (1994). Applying systems theory to systemic change: a generic model for educational reform. Paper presented the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 1994.
- Karpenko, O., Bershadskaja, M., & Voznesenskaia, I. (2009). The PISA international study and problems of the development of Higher Education. *Russian Education & Society*; 51(4), 68–91. doi:10.2753/RES1060-9393510404.
- Katz, D., & Kahn, R. L. (1966). *The social psychology of organizations*. New York: Wiley.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13, 3–9.
- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301–319). New York: McGraw-Hill.
- Kirkpatrick, D. L. (1996). Invited reaction: reaction to Holton article. *Human Resource Development Quarterly*, 7, 23–25.
- Landy, F. J., & Conte, J. M. (2007). *Work in the 21 century. An introduction to industrial and organizational psychology* (2nd ed.). Malden: Blackwell.
- Russell, T. R. (2004). Assessing outcomes of character-building programs: problems and prospects. *New Directions for Institutional Research*, 122, p105–p110. doi:10.1002/ir.113.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: a decade of progress. *Annual Review of Psychology*, 52, 471–499.
- Sax, L. J. (2004). Citizenship development and the American college student. *New Directions for Institutional Research*, 122, 65–80. doi:10.1002/ir.110.
- Toutkoushian, R. K. (2005). What can institutional research do to help colleges meet the workforce needs of states and nations? *Research in Higher Education*, 46, 955–984. doi:10.1007/s11162-005-6935-5.
- Van Buren, M. E., & Erskine, W. (2002). *The 2002 ASTD state of the industry report*. Alexandria: American Society of Training and Development.
- Voorhees, R. A., & Harvey, L. (2005). Higher education and workforce development: a strategic role for institutional research. *New Directions for Institutional Research*, 128, 5–12.