

Competency Assessment Toolkit for Professional Psychology

Nadine J. Kaslow
Emory University School of Medicine

Catherine L. Grus
American Psychological Association Education Directorate

Linda F. Campbell
University of Georgia

Nadya A. Fouad
University of Wisconsin-Milwaukee

Robert L. Hatcher
University of Michigan

Emil R. Rodolfa
University of California, Davis

A "toolkit" for professional psychology to assess student and practitioner competence is presented. This toolkit builds on a growing and long history of competency initiatives in professional psychology, as well as those in other health care disciplines. Each tool is a specific method to assess competence, appropriate to professional psychology. The methods are defined and described; information is presented about their best use, psychometrics, strengths and challenges; and future directions are outlined. Finally, the implications of professional psychology's current shift to a "culture of competency," including the challenges to implementing ongoing competency assessment, are discussed.

Keywords: competencies, assessment, benchmark

THIS ARTICLE REFLECTS the efforts of a Task Force Group supported by the Board of Educational Affairs of the Education Directorate of the American Psychological Association. The first two authors were the co-chairs of the Task Force and took the lead in writing the manuscript. The remaining authors actively participated in the Task Force and the crafting of the manuscript and are listed in alphabetical order. Nadine J. Kaslow, Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia; Catherine L. Grus, Education Directorate, American Psychological Association, Washington, DC; Linda F. Campbell, Department of Counseling and Human Development Services, University of Georgia, Athens, Georgia; Nadya A. Fouad, Department of Educational Psychology, University of Wisconsin-Milwaukee. Robert L. Hatcher, Department of Psychology, University of Michigan, Ann Arbor, Michigan; Emil R. Rodolfa, CAPS-Counseling and Psychological Services, University of California, Davis.

THE AUTHORS OF THIS ARTICLE were the members of the American Psychological Association's workgroup on the Competency Assessment Toolkit in Professional Psychology. The lead authors took primary responsibility for crafting this article. The first author was the Task Force Chair. The remaining authors are listed in alphabetical order.

NADINE J. KASLOW, PhD, earned her doctorate in clinical psychology from the University of Houston. She is Professor and Chief Psychologist at Emory University School of Medicine, Department of Psychiatry and Behavioral Sciences at Grady Hospital and Special Assistant to the Provost. Currently, she is President of Division 29 and of the American Board of Clinical Psychology and Incoming Editor of the *Journal of Family Psychology*. She was the recipient of the APA Distinguished Contributions to Education and Training Award (2004). Her research and clinical interests focus on competency-based education, training, and supervision of interns and postdoctoral fellows; family violence; suicidal behavior across the life-span; and family systems medicine.

CATHERINE L. GRUS, PhD, is the Associate Executive Director for Professional Education and Training at the American Psychological Association (APA). Dr. Grus received her PhD in clinical psychology from Nova University. At APA, Dr. Grus works to advance policies and practices that promote quality education and training in professional psychology.

LINDA F. CAMPBELL, PhD, received her doctorate in counseling psychology from Georgia State University. She is Professor and Director of the Center for Counseling at the University of Georgia. Her research focuses on ethics, supervision, and psychotherapy.

NADYA A. FOUAD, PhD, is professor and training director of the Counseling Psychology program at the University of Wisconsin-Milwaukee. She received her doctorate from the University of Minnesota. She was recipient in 2003 of the John Holland Award for Outstanding Achievement in Career and Personality Research. She is Editor of *The Counseling Psychologist*. She has published articles and chapters on cross-cultural vocational assessment, career development of women and racial/ethnic minorities, interest measurement, cross-cultural counseling, and race and ethnicity.

ROBERT L. HATCHER received his PhD in Clinical Psychology from the University of Michigan, where he is currently the director of the Psychological Clinic. He is president emeritus of the Association of Directors of Psychology Training Clinics. His research interests include the alliance in therapy, interpersonal measurement, and professional competencies.

EMIL RODOLFA received his PhD from Texas A&M University. He is the director of Counseling and Psychological Services at the University of California, Davis. Currently, he serves as the President of the Association of State and Provincial Psychology Boards (ASPPB) and is a member of the State of California Board of Psychology. He is the Editor of *Training and Education in Professional Psychology*. His interests include legal and ethical practice and supervision and training.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Nadine J. Kaslow, PhD, Emory University School of Medicine, Department of Psychiatry and Behavioral Sciences, Grady Hospital, 80 Jesse Hill Jr. Drive, Atlanta, GA 30303. E-mail: nkaslow@emory.edu

Background of the Assessment of Competence Movement

Competence as the outcome of education and training is valued and increasingly addressed by health care professions (Bandiera, Sherbino, & Frank, 2006; Cowan, Norman, & Coopama, 2005; Kak, Burkhalter, & Cooper, 2001; Kaslow, 2004; Kaslow & Bell, 2008; Kaslow et al., 2004; Medical School Objectives Writing Group, 1999; Nelson, 2007; Spielman, Fulmer, Eisenberg, & Alfano, 2005; Watson, Stimpson, Topping, & Porock, 2002). This approach is a contrast with earlier models of education and training that focused on course objectives and a curriculum that was designed to meet those objectives (Nelson, 2007).

Professional psychology is currently shifting to a "culture of competence" (Roberts, Borden, Christiansen, & Lopez, 2005). Competence in health care providers is demanded by consumers, expected and certified by regulators, and lauded by policymakers (Hoge et al., 2005). The assessment of competence offers much to education and training; it fosters learning, evaluates progress, assists in determining curriculum and training program effectiveness, advances the field, and protects the public (Kaslow, 2004).

Although efforts to define and measure competence have burgeoned over the past few years, notable initiatives date back several decades (Nelson, 2007; Rubin et al., 2007). The first widely disseminated model that articulated competencies for education and training programs was developed by the National Council of Schools and Programs of Professional Psychology (NCSPP; Peterson, Peterson, Abrams, & Stricker, 1997). Other models emerged through the work of the Council of Chairs of Training Councils (CCTC)/Association of Directors of Psychology Training Clinics (ADPTC) Practicum Competencies in 2001 (Hatcher & Lassiter, 2007). The major initiative that pushed the competency agenda to the forefront was the Competencies Conference: Future Directions in Education and Credentialing (Kaslow, 2004; Kaslow et al., 2004). At this conference, there was agreement on the core foundational (Bieschke, Fouad, Collins, & Halonen, 2004; Daniel, Roysircar, Abeles, & Boyd, 2004; de las Fuentes, Willmuth, & Yarrow, 2005; Elman, Illfelder-Kaye, & Robiner, 2005) and functional competencies (Arredondo, Shealy, Neale, & Winfrey, 2004; Falender et al., 2004; Krishnamurthy et al., 2004; Spruill et al., 2004), methods for education and training related to these competencies, and strategies for assessing the competencies (Roberts et al., 2005). A "cube model" captured the intersection of the foundational and functional competencies in a fashion that takes into account developmental stage (Rodolfa et al., 2005).

The cube model was further operationalized by the Assessment of Competency Benchmarks Workgroup and described in this issue (Fouad et al., 2009), the product of that group serves as the basis for this toolkit. This group further articulated the overarching core competencies based on advances in the field since the Competencies Conference. The currently agreed upon foundational domains of competence include professionalism, reflective practice, scientific knowledge and methods, relationships, individual and cultural diversity, ethical and legal standards and policy, and interdisciplinary systems. The functional competencies include the following: assessment, intervention, consultation, research/evaluation, supervision, teaching, administration, and advocacy. This workgroup identified essential components of each benchmark and behavioral anchors for the foundational and functional competencies over the formal education and training sequence.

The APA Board of Educational Affairs (BEA) Task Force on Assessment of Competence in Professional Psychology (2004) provided a seminal report (Kaslow, Rubin, Bebeau, et al., 2007), as well as several publications that have informed the training community's thinking about methods to assess competence. This Task Force focused on the history of the competencies movement within professional psychology (Rubin et al., 2007), assessment models from other health professions (Leigh et al., 2007), and challenges in the assessment of competence (Lichtenberg et al., 2007), and offered guiding principles for the assessment of competence within professional psychology (Kaslow, Rubin, Bebeau, et al., 2007). This article will present the development of the toolkit; definitions related to the toolkit; descriptions, psychometric properties, uses and limitations of several key types of assessment in professional psychology; and finally, Table 1 provides useful types of assessment tools for each competence and its components. Table 2 presents usefulness ratings of Toolkit assessment methods to measure specific essential elements.

Background on the Competency Assessment Toolkit for Professional Psychology

The Competency Assessment Toolkit for Professional Psychology was developed to respond to a need in the education and training and regulatory communities to provide best practice resources for graduate, internship, and postdoctoral programs and licensing boards wishing to systemically develop and assess the acquisition of competence. The competency assessment workgroup based much of our work on the Accreditation Council for Graduate Medical Education and American Board of Medical Specialties (ACGME/ABMS) Toolbox of Assessment Methods that stimulated the development of this product (Accreditation Council for Graduate Medical Education and American Board of Medical Specialties, 2000), as well as the efforts of our psychiatric colleagues related to defining, teaching, and assessing the competencies focal to psychiatric education (Andrews & Burruss, 2004; Dingle & Beresin, 2006; Swick, Hall, & Beresin, 2006). Although the information presented reflects the current state of the art within professional psychology, the knowledge base and available assessment armamentarium will expand and these materials will need to be updated periodically. The tools lend themselves best to programs preparing students for health service provision careers and regulatory bodies that credential professionals in the health service arena and while there may be broader applications, these were the targeted audience. The intent is not to create a culture of assessment, but to refine the methods used by training programs and credentialers so that the assessment methods used to ascertain individuals' progression through the educational, training, and credentialing sequence are optimal for tapping the learning outcomes of greatest relevance to effective professional functioning.

Definitions

The following definitions are offered to guide the user. Historically, a number of terms that describe overarching concepts have been used interchangeably, when in fact they refer to discrete constructs (Cowan et al., 2005). Competence is "the habitual and judicious use of communication, knowledge, technical skills, clinical

(text continues on page S34)

all
med
profs

Table 1
Toolkit Assessment Measures, Foundational and Functional Competencies, Formative and Summative Evaluation, and Developmental Level

Method	Overall broad competencies useful	Essential components of competencies useful	Formative and summative evaluation (notes types of evaluation most useful for most competencies)	Developmental level (notes levels most useful for)
360-degree evaluation	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Reflective practice ● Relationships ● Ethical and legal standards and policy ● Interdisciplinary systems ● Functional ● Supervision ● Teaching ● Administration ● Advocacy 	<ul style="list-style-type: none"> ● Foundational ● Individual and cultural diversity ● Functional ● Intervention 	Formative Summative	Readiness for internship Entry level to practice Advanced credentialing
Annual/rotation Performance Reviews	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Relationships ● Individual and cultural diversity ● Ethical and legal standards and policy ● Interdisciplinary systems ● Functional ● Supervision 	<ul style="list-style-type: none"> ● Foundational ● Reflective practice ● Scientific knowledge and methods ● Functional ● Assessment ● Intervention ● Research and evaluation ● Administration ● Advocacy 	Summative	Readiness for practicum Readiness for internship Entry level to practice Advanced credentialing
Case presentation reviews	<ul style="list-style-type: none"> ● Foundational ● Ethical and legal standards and policy ● Functional ● Intervention 	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Reflective practice ● Scientific knowledge and methods ● Relationships ● Individual and cultural diversity ● Interdisciplinary systems ● Functional ● Assessment ● Supervision ● Consultation ● Research and evaluations 	Formative Summative	Readiness for internship Entry level to practice Advanced credentialing
Competency Evaluation Rating Forms (CERF)	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Relationships ● Individual and cultural diversity ● Ethical and legal standards and policy ● Functional ● Intervention ● Supervision 	<ul style="list-style-type: none"> ● Foundational ● Scientific knowledge and methods ● Interdisciplinary systems ● Functional ● Assessment ● Research and evaluation ● Teaching 	Formative Summative	Readiness for practicum Readiness for internship Entry level to practice Advanced credentialing
Client/patient process and outcome data	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Relationships ● Individual and cultural diversity 	<ul style="list-style-type: none"> ● Foundational ● Ethical and legal standards and policy ● Functional ● Assessment ● Intervention ● Consultation 	Formative	Readiness for internship Entry level to practice Advanced credentialing

(table continues)

Table 1 (continued)

Method	Overall broad competencies useful	Essential components of competencies useful	Formative and summative evaluation (notes types of evaluation most useful for most competencies)	Developmental level (notes levels most useful for)
Consumer surveys	<ul style="list-style-type: none"> Foundational ● Professionalism ● Relationships ● Individual and cultural diversity Functional ● Supervision ● Administration ● Advocacy 	<ul style="list-style-type: none"> Foundational ● Scientific knowledge and methods ● Ethical and legal standards and policy ● Interdisciplinary systems Functional ● Assessment ● Intervention ● Consultation ● Teaching 	Formative Summative	Readiness for internship Entry level to practice Advanced credentialing
Live or recorded Performance ratings	<ul style="list-style-type: none"> Foundational ● Relationships ● Individual and cultural diversity ● Interdisciplinary systems Functional ● Consultation ● Administration 	<ul style="list-style-type: none"> Foundational ● Professionalism ● Scientific knowledge and methods ● Ethical and legal standards and policy Functional ● Assessment ● Intervention ● Research and evaluation ● Supervision ● Teaching 	Formative Summative	Readiness for practicum Readiness for internship Entry level to practice Advanced credentialing
Objective Structured Clinical Examinations (OSCEs)	<ul style="list-style-type: none"> Foundational ● Relationships ● Ethical and legal standards and policy Functional ● Assessment ● Intervention 	<ul style="list-style-type: none"> Foundational ● Professionalism ● Reflective practice ● Scientific knowledge and methods ● Individual and cultural diversity Functional ● Consultation ● Research and evaluation 	Summative	Readiness for internship Entry level to practice Advanced credentialing
Portfolios	<ul style="list-style-type: none"> Foundational ● Scientific knowledge and methods Functional ● Intervention ● Research and evaluation ● Teaching ● Administration 	<ul style="list-style-type: none"> Foundational ● Professionalism ● Reflective practice Functional ● Assessment ● Supervision 	Summative Formative	Readiness for practicum Readiness for internship Entry level to practice Advanced credentialing
Record reviews	<ul style="list-style-type: none"> Foundational ● Ethical and legal standards and policy 	<ul style="list-style-type: none"> Foundational ● Professionalism ● Ethical and legal standards and policy Functional ● Assessment ● Intervention ● Consultation ● Supervision 	Formative	Readiness for internship Entry level to practice Advanced credentialing
Simulations/role plays	<ul style="list-style-type: none"> Foundational ● Ethical and legal standards and policy 	<ul style="list-style-type: none"> Foundational ● Relationships ● Functional ● Assessment ● Intervention 	Formative	Readiness for practicum Readiness for internship Entry level to practice Advanced credentialing

(table continues)

Table 1 (continued)

Method	Overall broad competencies useful	Essential components of competencies useful	Formative and summative evaluation (notes types of evaluation most useful for most competencies)	Developmental level (notes levels most useful for)
Self-assessment	<ul style="list-style-type: none"> ● Foundational ● Reflective practice ● Individual and cultural diversity ● Ethical and legal standards and policy ● Interdisciplinary systems ● Functional ● Supervision ● Administration ● Advocacy 	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Scientific knowledge and methods ● Relationships ● Functional ● Assessment ● Intervention ● Consultation ● Research and evaluation ● Teaching 	Formative	<ul style="list-style-type: none"> ● Readiness for practicum ● Readiness for internship ● Entry level to practice ● Advanced credentialing
Standardized client/patient interviews	<ul style="list-style-type: none"> ● Foundational ● Ethical and legal standards and policy 	<ul style="list-style-type: none"> ● Foundational ● Scientific knowledge and methods ● Relationships ● Individual and cultural diversity ● Functional ● Assessment ● Intervention ● Research and evaluation ● Teaching 	Summative	<ul style="list-style-type: none"> ● Readiness for internship ● Entry level to practice ● Advanced credentialing
Structured oral examinations	<ul style="list-style-type: none"> ● Foundational ● Scientific knowledge and methods ● Ethical and legal standards and policy 	<ul style="list-style-type: none"> ● Foundational ● Professionalism ● Reflective practice ● Individual and cultural diversity ● Interdisciplinary systems ● Functional ● Assessment ● Intervention ● Consultation ● Supervision 	Summative	<ul style="list-style-type: none"> ● Readiness for practicum ● Readiness for internship ● Entry level to practice ● Advanced credentialing
Written examinations	<ul style="list-style-type: none"> ● Foundational ● Scientific knowledge and methods ● Ethical and legal standards and policy 	<ul style="list-style-type: none"> ● Foundational ● Individual and cultural diversity ● Interdisciplinary systems ● Functional ● Assessment ● Intervention ● Consultation ● Research and evaluation ● Teaching 	Summative	<ul style="list-style-type: none"> ● Readiness for practicum ● Readiness for internship ● Entry level to practice

Table 2
Usefulness Ratings of Toolkit Assessment Methods to Measure Specific Essential Elements

Competency	Essential element	Assessment method					
		Annual review	Case reviews	Process/outcome data	Rating forms	Consumer surveys	Objective structured clinical examinations
Professionalism	Integrity	1	3	1	1	2	
	Depotment	1	2	1	1	1	2
	Accountability	1	1	1	1	1	
	Concern welfare of others	1	1	1	1	1	2
Reflective practice	Professional identity	1	2	3	1	2	2
	Reflective practice	1	2		2		2
	Self-assessment		2		2		
Scientific knowledge and methods	Self-care	2		3	2		
	Scientific mindedness	2	2		2		3
Relationships	Scientific foundation of psychology	2	3		2		1
	Scientific foundation practice	2	1	2	2	1	1
	Interpersonal relationships	1	2	1	1	1	1
	Affective skills	1	2	1	1	1	2
ICD	Expressive skills	1	1	1	1	1	1
	Self shaped ICD and context	1	2	3	1		3
	Others shaped by ICD and context	1	1	1	1	1	2
Ethical legal standards policy	Interaction self and others shaped by ICD	1	3	1	2	1	1
	Apps. based on individual and cultural context	1	1	1	1	1	2
	Ethical, legal, and professional standards and guidelines	1	1		2		2
	Awareness and application of ethical decision making	1	1	2	1	2	1
Interdisciplinary systems	Ethical conduct	1	3	1	1	1	1
	Knowledge shared/distinctive contributions of disciplines	1	3		2	3	
	Functioning in multi- and inter-disciplinary contexts	1	3	3	2	1	
	Participation in interdisciplinary enhances outcomes	3	1		3		
Assessment	Respectful and productive relationship with other disciplines	1	1	3	1	1	2
	Measurement and psychometrics	2	1		2		
	Evaluation methods	2			2	3	
	Application methods	2		3	2	2	1
Intervention	Diagnosis	2	1	2	2	1	1
	Conceptualization and recs	2	1	2	2	1	2
	Communication findings	2		1	2	1	1
	Knowledge interventions	2	3	3	2		
Consultation	Intervention planning	2	1	2	1	2	2
	Skills	2	2	1	3	2	1
	Implementation	1	1	1	1	1	1
	Progress evaluation	3	1	2	2	2	
Research and evaluation	Role of consultant	3	2	2	2	3	
	Address referral question	2	1	2	2	1	1
	Communication findings	3		1	2	1	1
Supervision	Application methods	2	1	1	2	1	1
	Scientific approach to knowledge generation	1			2		
Teaching	Application scientific method to practice	2	1	2	2	1	1
	Expectations and roles	1			1	2	
	Processes and procedures	1	3		1	1	1
	Skills development	2	3		2	1	1
Administration	Aware factors affect quality	2	3		2	2	
	Participation process	2	1		2	1	
	Ethical and legal issues	2	1			2	
Advocacy	Knowledge				1		
	Skills	2			1	1	
Administration	Management	2	3		2	1	
	Administration	2	3		2	1	
Advocacy	Leadership	1			2	1	
	Empowerment	1	2		2	1	
	Systems change	2	2		2	1	

Note. ICD = Individual and cultural diversity. 1 = very useful method; 2 = useful; 3 = potentially useful.

Assessment method								
Portfolio reviews	Ratings performance	Record reviews	Self-assessment	Simulations/role plays	Standard patient interviews	Standard oral exams	Written exams	360-degree evaluations
2		3	3			3		1
3	1		2	2	2	1		1
2		1	2			2		1
	1		1	3	2	2		1
1			1			1		1
1	2		1	3	3	1		1
2			1	3	3	2		1
			1					1
2	2		3	2	2	1	1	
2	3		3	3	3	2	1	
1	1	3	2	2	1	1	3	
3	1		1	1	1	2		1
3	1		1	2	2	2		1
1	1	2	2	2	2	2	1	1
2	2		1	3	3	1	3	2
2	1	3	1	2	2	2	3	2
2	1		1	2	2	2		1
1	1	3	2	2	2	2	3	
2	2	3	1	2	2	1	1	
2	2	2	1	1	1	1	1	
2	1	2	1	1	1	2	3	1
		3				1	1	
2	1		1			2		1
3	3			3		1	3	
3	1	3	1					1
2		3				2	1	
3		3		1		2	1	
2	1	2	2	2	2	2	3	
2	3	2	2	2	2	1	1	
2		1	2	2	2	1	2	
1	1		2	1	2	2		2
			2			1	1	
2	3	1		2	2	1		
2	1	2	1	1	1	2		
2	1	2	1	2	2	2	3	1
3	2	2	1			2		2
3			3			1	1	
3	1	2		3	2	2		
3	1	1	2	3	2	2		2
2	1	3	1	3	2	2	3	2
1								
2	1		2	2	1	2	2	
3	1	3	1	3	2	1	2	1
2	1	2	2	1	1	2	3	1
3			1			1		
2	1		2			2	3	
2	3	1	2			1	2	
1			2	2	3	1	1	
2	1	2	1	2	1	2		1
2		2	1			3	3	1
2		2	1			3	3	1
1			1			3	3	1
2	2		1			3	3	1
2			1					1
2			1			3		1

reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" (Epstein & Hundert, 2002, p. 227). Professional competence consists of cognitive, integrative, relational, affective/moral, and habits of the mind dimensions. It is developmental and context-dependent. Competencies are demonstrable elements or components of performance (knowledge, skills, and attitudes and their integration) that make up competence. Foundational competencies refer to the knowledge, skills, attitudes, and values that serve as the foundation for the functions a psychologist is expected to carry out and are cross-cutting, relevant to each of the foundational and functional competencies, whereas functional competencies encompass the major functions that a psychologist is expected to carry out (Rodolfa et al., 2005). Formative evaluations assess competence and provide ongoing corrective, developmentally informed feedback to the individual to foster growth (Black & William, 1998). Summative evaluations measure outcomes at the end-point of a developmental process for purposes of progression and gatekeeping (Black & William, 1998).

Terms specific to assessment methodology deserve note. Reliability is the degree to which an assessment will yield similar results if repeatedly administered under comparable conditions and that scores for a given individual will be consistent when assessed by different methods, by different raters, or for more than one encounter. How well the assessment measures the competencies it intends to measure is the definition of validity. Fidelity refers to the extent to which the method of assessment approximates the actual behavior that is being measured.

The following terms pertain to the product of the Benchmarks Workgroup (Fouad et al., 2009). Benchmarks are the behavioral indicators associated with each domain that provide descriptions and examples of expected performance at each developmental stage, and are standards for measurement of performance that can be used for comparison, to identify where needs for improvement exist, and to determine if a given competency has been achieved (Carraccio & Englander, 2004). Essential components are the critical elements of the knowledge/skills/attitudes that make up the particular competency. Behavioral anchors describe what essential components would look like if observed, using operational terms. Developmental levels refer to stages of professional development. The Benchmarks document focuses on three developmental transitions: readiness for practicum, readiness for internship, readiness for entry level to practice.

Tools

This section delineates, in alphabetical order, the tools we recommend for inclusion in a comprehensive assessment armamentarium for professional psychology. For each tool, we (1) offer a description, (2) detail implementation procedures, (3) review psychometric properties, (4) highlight key strengths, and (5) acknowledge salient challenges. Table 1 lists each assessment method and addresses its application to overall or broad competency domains and essential components of these domains, predominant use for formative and summative evaluation, and developmental levels for which it is most appropriate. Table 2 presents in more detail information about how useful each tool is for assessing the essential elements/subcomponents of each competency; 1 = *very useful*, 2 = *useful*, 3 = *potentially useful*, and no

number denotes *not indicated for use*. This document does not describe specific instruments to be used, but rather provides overarching comments regarding each methodology. To a large extent, the lack of specific instruments cited reflects the current state of the art in the assessment of competence in professional psychology. Similarly, the general comments made about psychometrics, which are not tied to specific tools, is a further reflection of the status of our assessment efforts.

The Competency Assessment Toolkit for Professional Psychology was designed to be a companion to the Assessment of Competency Benchmarks Workgroup in the following ways (Fouad et al., 2009). First, the toolkit delineates appropriate methods for assessing each of the overall and broad foundational and functional competencies outlined in the Benchmarks document. Second, the toolkit outlines relevant assessment strategies for measuring the essential components of each of the foundational and functional competencies. Third, the toolkit discusses the appropriateness of each tool for measuring competency at the three levels of education and training that are the focus of the Benchmarks document, while adding a fourth level of professional development (advanced credentialing).

360-Degree Evaluations

Description. The 360-degree evaluations glean systematic input retrospectively, concurrently, and individually from multiple raters in the person being assessed's sphere of influence regarding key performance behaviors and attitudes (Lockyer, 2003). Others have advocated for the use of this assessment approach across a variety of foundational and functional competency domains (Fletcher & Bailey, 2003; Joshi, Ling, & Jaeger, 2004; Manring, Beitman, & Dewan, 2003; Rodgers & Manifold, 2002; Sidhu, Grober, Musselman, & Reznick, 2004).

Implementation. The first step for implementing 360-degree evaluations is choosing the actual instrument. Then, it is important to ascertain who will serve as raters. Raters typically include supervisors, a diverse cadre of peers and colleagues including those from other disciplines, subordinates (e.g., supervisees), and the person being assessed, and may include the clients/patients of the person being assessed. Once raters are chosen, the person being assessed invites them to serve as raters. After an orientation to the process, raters complete the comprehensive evaluation using paper-based measurement tools (surveys, questionnaires) or online via the use of computer software packages and use the rating scales to assess how frequently and effectively a behavior is performed or an attitude is observed and how important the behavior or attitude is to the context. When the assessment tool offers the option, raters add comments illustrating the reasons for the specific ratings. Once the ratings are obtained, they are summed across all evaluators by topic or competency. Then a trained person, typically someone who receives intensive instruction by an organization that specializes in this assessment method, provides detailed feedback to the person being assessed and discusses the similarities and differences of ratings across informants and areas to target for growth. Developing, with the aid of a trained professional, an action plan to address areas for self-improvement, is the final step in the implementation process.

Psychometrics. There is significant empirical support for the psychometrics of 360-degree evaluations in leadership and busi-

ness contexts (Atkins & Wood, 2002), including high levels of internal consistency and interrater reliability and initial evidence that 360-degree evaluation data correlate with other types of ratings. With health professionals outside of psychology, there is some support for the construct and convergent validity and interrater reliability of this method (Lockyer, 2003).

Strengths. 360-degree evaluations, one of the best methods for assessing the breadth of foundational competencies, offer fair, accurate, objective, and well-rounded assessments of the person being assessed and allow the person being assessed to gain a greater appreciation of how they are viewed by others, areas of strength, aspects of personal functioning that can be improved on, and where there are discrepancies between self-perceptions and the perceptions of others (Fletcher & Bailey, 2003). Engaging in the 360-degree evaluation processes bolsters understanding about the competency framework relevant to the organization or program. Including 360-degree evaluations in an organization offers a culture shift that values the provision and receipt of feedback, as long as feedback is given in accord with best practices (Carson, 2006).

Challenges. This assessment method is associated with the following challenges. There often are difficulties in constructing a survey appropriate for use by all evaluators in the circle of influence, which may require the triangulation and integration of different assessments from different informants (Manring et al., 2003). Orchestrating data collection from a large number of individuals is no easy feat (Manring et al., 2003). Evaluators often are concerned about the confidentiality of their feedback, given its sensitive and detailed nature. There are questions regarding the reliability and validity of feedback from certain raters and the appropriateness of gathering data from some informants (e.g., clients/patients). Misuse of 360-degree feedback may be associated with anxiety and hurt feelings, which might negatively impact performance (Carson, 2006). This approach entails significant costs and resources and it is unclear if the incremental benefits outweigh the resource costs and work involved (Weigelt, Brasel, Bragg, & Simpson, 2004).

Annual/Rotation Performance Reviews

Description. Annual/rotation performance reviews frequently are conducted in professional psychology, but little has been written on this assessment method. These annual or end of rotation reviews entail faculty, supervisors, and possibly peers evaluating the foundational and functional competencies of the person being assessed and the multisource feedback is integrated into a comprehensive summative formulation (Epstein, 2007). Recently, some attention has been paid in medicine to psychometrically sound instruments for peer assessments that could be modified and incorporated into these annual/rotation performance reviews (Evans, Elwyn, & Edwards, 2004).

Implementation. The first step in implementing these reviews is to identify the competencies to be evaluated and the assessment sources that will comprise the review. Then, it is necessary to determine the rating method(s) and feedback mechanism. Input from the various assessors needs to be integrated and the performance of the person being assessed needs to be compared against the behavioral anchors for the given developmental level. Then, the person being assessed needs to receive specific feedback to target competencies and their essential components to enhance.

Psychometrics. The limited psychometric analyses related to this method provide some evidence that assessment from multiple viewpoints increases construct validity and that direct observation may increase validity and reliability (Kak et al., 2001). The more global the assessments are and the less complex the skills being rated, the greater the agreement between informants (Falchikov & Goldfinch, 2000). Consideration needs to be given to the potential for biases to affect ratings (e.g., halo effect).

Strengths. Annual/rotation performance evaluations provide an easy to use and inexpensive method for competency assessment. They offer the opportunity to utilize assessment of essential components to yield global ratings. Furthermore, they allow for more encompassing evaluations that include foundational competencies (e.g., professionalism) in addition to functional competencies.

Challenges. This approach requires time to evaluate all students or trainees in a program and provide them with meaningful feedback, which may be prohibitive resulting in students or trainees receiving general and nonspecific feedback. The global nature of the assessment and the fact that it frequently does not entail direct observation often limits the detail that faculty members or supervisors provide to the person being assessed (Epstein, 2007; Kak et al., 2001). Furthermore, different concerns influence assessors' ratings of skills and behaviors; peers often focus on relationships, supervisors tend to overemphasize functional competencies that are client/patient focused, and faculty may over rely on academic information and those competencies most related to scholarship. Assessments may be influenced by the assessors' relationships with and views of the person being assessed. For such evaluations to be effective, training for consistency across assessors is required. For assessors to provide meaningful and integrative feedback in a fashion that takes into account the anxiety that the person being assessed may experience in receiving summative feedback.

Case Presentation Reviews

Description. Case presentation reviews are common practice within professional psychology, although they often are viewed more as a teaching/supervisory method than as a formal assessment of competence. In the case presentation review, the person being assessed discusses client/patient/system characteristics, assessment methods, intervention planning, implementation, and outcome (Petti, 2008). Assessors evaluate the case presentation and the person being assessed's understanding of the client/patient/system, application of theory and evidence base, implementation efforts, and personal reactions.

Implementation. Implementation begins with identifying the competencies to be assessed for all parties. Then, the person (people) conducting the case review should provide the person to be assessed a framework to present and discuss in writing and/or verbally the case using the following categories: client/patient/system background information; presenting problem; history; mental status; assessment; conceptualization; intervention plan and implementation; future plans; and references. The assessor(s) need to be trained in any rating scale that may be used. During case reviews, the person presents a case for a specific amount of time, followed by an interactive dialogue with the assessor(s). It is useful to combine this method with live and recorded performance.

Psychometrics. Despite the popularity and common use of this method, there is limited psychometric information available, because rating case presentations is typically informal and not standardized or when it is standardized, the reliability and validity has not been studied. Recently, a formal process to provide a summative evaluation of clinical competencies by evaluating specific aspects of a case presentation has been described and this approach yields adequate overall reliability of case reviews and offers training programs a normative data set (Petti, 2008).

Strengths. Case review presentations enable assessors to hear the person being assessed describe knowledge application, skills, and values during interactions with clients/patients/systems. They provide a method to evaluate verbal and nonverbal communication. This approach offers a familiar method in most contexts. Case review presentations give assessors and systems a low cost, low resource intensive, and feasible method.

Challenges. This approach raises questions about the accuracy of recall on the part of the person being assessed. It requires effective written and oral communication. Case review presentations elicit concerns by the person being assessed about sharing details of interactions with clients/patients/systems or reflective practice.

Client/Patient Process and Outcome Data

Description. Client/patient process and outcome data may be gleaned from measures of the therapeutic or working alliance, self-report symptom checklists, or ratings from the therapist/assessor/consultant or an independent assessor. Working alliance measures, such as the Working Alliance Inventory (Horvath & Greenberg, 1986) and the short form of this measure (Hatcher & Gillaspay, 2006) assess the quality of the working relationship between the assessor/therapist/consultant and client/patient and are indicators of process. Symptom checklists, rating scales, and diagnostic interviews assess subjective distress, psychiatric symptoms, degree of impairment in life functioning, strengths, and progress and can be used as markers of outcome (Mariush, 2004). Others have asserted that this approach can be useful in assessing the overall and essential components of some foundational and functional competencies (Manring et al., 2003).

Implementation. This approach can be implemented by initially determining what measure(s) to use and the timing of the assessments. Then, the purpose(s) of the data gathering needs to be explained to the client/patient. Finally, the process and outcome data are used to enhance the quality of the services rendered.

Psychometrics. A considerable body of evidence reveals moderate to high reliability and validity of working alliance measures, symptom checklists, clinician ratings, and structured diagnostic interviews.

Strengths. These measures provide a relatively inexpensive and easy to administer and score way to assess client/patient process and outcome, one marker of performance on the part of the person being assessed. It includes the potential to incorporate widely used, standardized, reliable and valid instruments that assess global functioning or discrete symptoms/disorders.

Challenges. Such measures may elicit resistance on the part of the client/patient with regard to participation and may raise questions about the level of distortion in clients/patients' responses to surveys due to social desirability, reactivity, and transference,

which in turn presents challenges in interpreting the ratings. Also, the use of such measures to assess competence must consider that clients/patients are not assigned randomly to providers and that the quality of care that clients/patients receive in health care organizations is not due solely to the individual provider (Landon, Normand, Blumenthal, & Daley, 2003).

Competency Evaluation Rating Forms (CERF)

Description. CERFs are written documents that consist of a list of the behavioral indicators for selected foundational and functional competencies. They involve rating an individual on each behavioral indicator according to a numerical system that corresponds with levels of competence attainment. CERFs are a popular assessment tool in professional psychology.

Implementation. Prior to implementing the CERF, the competencies to be assessed must be identified and a Likert scale developed. Then, the CERF is completed and reviewed with the person being assessed (Bienenfeld, Klykylo, & Knapp, 2000).

Psychometrics. CERFs have high face, construct, content, and discriminant validity (Andrews & Burruss, 2004; Lievens & Sanchez, 2007). When observers are trained, moderate to good reliability can be achieved (Andrews & Burruss, 2004; Lievens & Sanchez, 2007). Without rater training on the definition of the competencies being evaluated, reliability of the CERF can be low across settings and across raters (Kak et al., 2001).

Strengths. CERFs provide an easy to use and inexpensive method for competency assessment (Hobgood, Riviello, Jouriles, & Hamilton, 2002). They enable raters to ascertain levels of competency acquisition on a continuum, which a dichotomous (pass/fail) rating does not allow. This strategy facilitates the tailoring to the specific behavioral indicators for the essential elements of selected competencies; pinpoints specific areas in need of improvement for a person being assessed; and serves as a useful assessment approach across the span of education and training, reflecting the development of various levels of competence longitudinally.

Challenges. CERFs pose difficulties for ensuring interrater reliability and requires direct observation data on which to base assessments, which may not always be available. They may not effectively assess the complex essential components of various competencies (Kak et al., 2001).

Consumer Surveys

Description. Self-report consumer surveys focus on satisfaction with the services rendered and assess the extent to which the services satisfy the desires and expectations of the client/patient/supervisee/student/employees and so forth. Whereas client/patient process and outcome data focuses on the progress made by the client/patient and his or her experience of the therapeutic relationship, consumer surveys are concerned with service delivery satisfaction. Consumer surveys address the consumer's perspective on the foundational and functional competencies of the person being assessed (Andrews & Burruss, 2004) and are useful to assess both overall and broad and essential components of a number of foundational and functional competencies (Chin, 2000; Manring et al., 2003).

Implementation. The use of consumer surveys as an assessment tool involves determining the competency(ies) to be assessed, mea-

sure(s) to be used, method(s) (e.g., phone, in person, mail), and timing of administration(s). There are some available measures for physicians (Evans, Edwards, Evans, Elwyn, & Elwyn, 2007) that could be adapted for use with psychologists. Data are collected from consumers and collated and then feedback is provided to the person being assessed. The reliability for most client/patient satisfaction surveys is moderate to high (Andrews & Burruss, 2004) and these tools have reasonable content validity (Dufrene, 2000). However, there is limited data on their construct validity or correlates with other attributes (Evans et al., 2007).

Strengths. Consumer surveys provide a relatively inexpensive and easy to administer and score assessment approach that can be used in a multitude of practice settings and offer valuable information from consumers regarding views about the relationship with the person being assessed and the quality of services.

Challenges. There are potential difficulties with the use of consumer surveys related to eliciting consumer participation; considering language and literacy problems; obtaining enough surveys per trainee to provide reproducible results; securing the resources required to collect, aggregate, and report survey responsibly; ascertaining the person being assessed's contribution to the consumer's care separate from that of the treatment team; interpreting the ratings because of questions about the level of distortion in consumers' responses to surveys associated with social desirability, reactivity, and transference; and ensuring that the feedback provided to the provider results in improved performance rather than professional resistance (Evans et al., 2007; Manring et al., 2003). Most consumer satisfaction surveys assess satisfaction with care, and how this translates into the competence of the provider is an empirical question (Grol, 2001).

Live or Recorded Performance Ratings

Description. This method consists of assessors directly observing live or via a recorded performance and then rating the person being assessed in accord with systematic competency criteria (behavioral indicators; Bandiera et al., 2006). A number of authors have advocated for the utilization of this assessment method for both overall broad foundational and functional competency domains, as well as the essential components of various domains (Jouriles, Burdick, & Hobgood, 2002; Jouriles, Emerman, & Cydulka, 2002; Manring et al., 2003; Tate, Foulkes, Neighbor, Campion, & Field, 1999).

Implementation. Implementation requires determining what format the observation will take, live (behind one-way mirror, in the room, Web cam) or recorded (audio, video). Once targeted competencies and their essential components to be evaluated are identified, the rating method to be used must be ascertained. Assessors must be trained to apply rating methods reliably in accord with systematic and specific criteria across people and performances. The method should be introduced to the person being assessed such that they become familiar with it, particularly if the method is newly introduced. It is necessary to orient the client/patient to the purpose and method of this approach and to secure informed consent from the client/patient.

Psychometrics. This method has good interrater reliability and content validity in medical settings (Ram et al., 1999; Tate et al., 1999).

Strengths. Live or recorded performance offers the assessor the opportunity to directly observe both foundational and functional competencies in which the person being assessed directly interacts with the client/patient. An excellent formative evaluation tool, they can be useful as a capstone performance demonstrating summative skills in an applied setting. This method encourages collaborative learning opportunities for improvement when assessors and the person being assessed observe and evaluate competencies together. It effectively tests verbal and nonverbal communication, offers an inexpensive logistically feasible methodology that is familiar in many training contexts, and provides an enduring record of the examinee's performance that facilitates evaluation at different times by different assessors.

Challenges. This approach may pose a challenge in terms of feasibility and practicality (e.g., difficulties with the video equipment and its operations) and although not expensive, it does involve costs (Ram et al., 1999). Ratings of live or recorded performance require informed consent, and can elicit resistance and other emotional reactions and concerns in the person being assessed and in the clients/patients being observed. Many clients/patients are resistant to being observed live or recorded, and as a result, this limits the sample for the assessment, raising questions about the generalizability of the findings from the evaluation. This method depends on assessors being trained to rate reliably (Holmboe, Hawkins, & Huot, 2004).

Objective Structured Clinical Examinations (OSCEs)

Description. OSCEs consist of several clinical encounters (called stations), typically with specially trained actors playing the role of a client/patient presenting with one or more psychological symptoms (Andrews & Burruss, 2004; Bandiera et al., 2006; Carraccio & Englander, 2000; Kak et al., 2001; Sidhu et al., 2004; Swick et al., 2006). Each encounter assesses a different competency or various essential elements of one or more competencies. OSCEs have been found to be excellent for summative assessments of both the essential components and overall or broad foundational and functional competencies (Colliver, Willis, Robbs, Cohen, & Swartz, 1998; Newble, 2004; Sidhu et al., 2004). Although this approach has empirical support for formative assessments (Adamo, 2003; Townsend, McIlvenny, Miller, & Dunn, 2001), using it in the context of such assessments may be too resource intensive and not feasible.

Implementation. Recently, a detailed review of creating the steps for an OSCE in psychiatry was published (Hodges, Hanson, McNaughton, & Regehr, 2002). While there are variants on the original OSCE format (Newble, 2004) key implementation steps are as follows. First, it is important to train the standardized clients/patients in the issues of the encounter and to ensure that they not provide any information unless the person being assessed requests it. Then, standardized client/patient encounters or tasks requiring interpretation of clinical information must be provided at separate stations that last for 5 to 10 minutes or longer and these must be observed by the assessor(s). At each station, the person being assessed completes notes or prepares other written material about the client/patient encounter. The person being assessed then moves between stations when a bell/buzzer/announcer indicates that it is time to move to the next station in accord with a specified sequence. Separate performance scores are generated for the tasks

at each station based on input from the standardized client/patient, the person being assessed and the assessor. The scores from multiple informants across the stations or tasks are combined by the assessors to reach a final outcome determination.

Psychometrics. With appropriate attention to design, OSCEs have acceptable psychometric properties when used in other health professions (Adamo, 2003; Carraccio & Englander, 2000; Sidhu et al., 2004; Swick et al., 2006), including good interrater, interstation, and split-half reliability; good generalizability; and strong content, construct, and concurrent validity. OSCE scores correlate moderately with other performance indicators and other forms of evaluations. The approach has increased reliability and content validity with a greater number of stations and similarity between tasks at different stations (Adamo, 2003; Carraccio & Englander, 2000). The method overall has a high degree of fidelity (Newble, 2004). No psychometric information can be located in the psychology literature on the OSCE competence assessment methodology.

Strengths. The literature documents a number of advantages of this assessment approach (Newble, 2004). The OSCE methodology measures clinical competence cross-sectionally using standardized means, focuses on observable behaviors, enables fairer peer comparison, and assesses complex competencies without endangering clients'/patients' well-being.

Challenges. OSCEs are challenging to create and administer given that they are labor intensive and costly and only cost-effective when many individuals are to be examined at one administration. They require high demands for other resources including large number of standardized clients/patients, significant number of assessors, and considerable time commitment for assessors (Kak et al., 2001; Swick et al., 2006). There are questions about the utility of this approach in evaluating some intervention essential components given that it does not provide a longitudinal assessment nor does it adequately tap complex skills requiring integrated professional judgment (Manring et al., 2003; Swick et al., 2006; Wass, van der Vleuten, Shatzer, & Jones, 2001).

Portfolios

Description. Portfolios are a collection of products, gathered by the person being assessed, which provide evidence of achievement of specific competencies (Fryer-Edwards, Pinsky, & Robins, 2006). They typically contain written documents, but also may include audio or video recordings or other forms of information. The content is not standardized and is implemented according to the desire of the program or credentialing body. The literature underscores the value of portfolios for assessing a few overall or broad and essential competencies of foundational and functional competencies (Carraccio & Englander, 2004; Fernsten & Fernsten, 2005; Fryer-Edwards et al., 2006; Lynch, Swing, Horowitz, Holt, & Messer, 2004). It has been found to be a strong tool for formative and summative evaluations across some competencies (Carraccio & Englander, 2004; Fernsten & Fernsten, 2005; Lynch et al., 2004; Manring et al., 2003; Swick et al., 2006; Wilkinson et al., 2002).

Implementation. Portfolio assessments entail deciding on the form (web-based or hard copy) and determining the elements (video tapes, assessment or treatment reports, evaluations). A mentoring system needs to be established. Efforts should be made

to facilitate assessor buy-in. It is essential that a supportive climate for learning and feedback be promoted. Outcomes and evaluation strategies must be determined (Carraccio & Englander, 2004; Fryer-Edwards et al., 2006).

Psychometrics. Reliability has not been well established because of the variable content included in a portfolio (Manring et al., 2003; Pitts, Coles, & Thomas, 1999; Swick et al., 2006) and the reliability and validity of the individual instruments included in the portfolio impact overall psychometric properties. Reproducible assessments are feasible when there is agreement on criteria and standards for the contents of the portfolio (Driessen, Van der Vleuten, Schuwirth, van Tartwijk, & Vermunt, 2005) and some evidence for construct and predictive validity has been established (Lynch et al., 2004).

Strengths. This relatively low cost assessment strategy has broad applicability. It allows for the assessment of actual work products and for items already generated for other purposes to be collected for the portfolio. It enables the person being assessed to share information about some activities and products that otherwise would have gone unnoticed. It expands over time as the person being assessed engages in additional activities such that more complex activities are increasingly reflected. It provides educational value and flexibility. This approach shifts responsibility for demonstrating competence to the person being assessed. Portfolio assessments serve as a tool for practice based learning and improvement that entails self-reflection and self-assessment in determining needs for improvement, developing a plan for attaining such, and measuring the effect of the plan in meeting goals. This methodology serves as a potentially useful tool to document continuing education activities (Wilkinson et al., 2002).

Challenges. The downsides of portfolios are that they require intense commitment of time and are labor intensive for all parties. Portfolio assessments require mentor involvement in the development and review of a portfolio. They may elicit resistance in the person being assessed. This approach evidences variable reliability/validity across the items evaluated in the portfolio (Carraccio & Englander, 2004; Fryer-Edwards et al., 2006).

Record Reviews

Description. Client/patient records (e.g., case files, psychotherapy notes, assessment protocols) developed and maintained by the person being assessed are reviewed by the assessor or administrative evaluators (Andrews & Burruss, 2004). Assessors employ record review to determine the presence, the quality of development, and the accuracy of essential elements of the client/patient case as recorded by the person being assessed. Standardized rating systems or protocols are developed to evaluate the critical elements of the record. Record reviews have been reported by others to be advantageous for assessing the essential components of a few foundational and functional competencies (Manring et al., 2003). They also have been found to be useful for formative evaluations for a limited to moderate number of competencies (Lynch et al., 2004).

Implementation. Criteria by which the records are assessed and are standardized or coded must be determined so that evaluation across records and across examinees may be reliable. Specific protocols for standard use to evaluate records systematically should be developed. Record reviews should be conducted with

scheduled frequency by assessors or administrative evaluators who have been trained to identify the expected elements of the record and to evaluate the standard of practice represented by the record.

Psychometrics. The use of standardized protocols and/or coding systems increases the reliability of evaluation. Ongoing case maintenance by an examinee provides greater stability in assessing performance. The development and maintenance of electronic clinical records avoids the problem of chart abstraction protocols. There is minimal empirical support for the reliability and predictive validity of this approach as applied to the assessment of competence (Lynch et al., 2004).

Strengths. Record reviews involve relatively low cost. They allow for the retroactive access to cases. This approach enables assessors to identify the accuracy and the descriptive recording of sentinel events, course of a case across time, decision-making, and follow through (Kak et al., 2001).

Challenges. Record reviews fail to capture all procedures, interventions, and treatment components, as they may not be recorded. They raise concerns about standardization, accuracy, and adherence to protocol recording. This method requires criteria to be set for coding records. It demands the conduct of multiple record reviews before feedback is effective (Kak et al., 2001; Manring et al., 2003).

Self-Assessment

Description. Self-assessment is the process by which the person being assessed validly ascertains personal and professional strengths and areas in need of improvement across foundational and functional competency domains, raises awareness of own limits of expertise and determines what to do when those limits are reached, and monitors own progress in the process of taking action to address specific developmental needs (Kaslow, Rubin, Forrest, et al., 2007). It has been considered by some to be an invaluable approach for evaluating specific essential elements of foundational and functional competencies, as well as overall and broad competencies (Belar et al., 2001; Pope, Sonne, & Greene, 2006).

Implementation. The person doing the self-assessment should learn the rationale for self-assessment. They should become familiar with self-assessment methodology. It is essential that they understand the competencies to be evaluated. It behooves them to become knowledgeable about the rating method to be used and delivery mechanism (e.g., paper, electronic). Finally, it is invaluable for them to reflect on results on their own or with the faculty member or supervisor responsible for the services being provided for which the self-assessment was performed. As a result of this examination, the self-assessment information should be used to guide the development of learning goals and action plans (Caverzagie, Shea, & Kogan, 2008).

Psychometrics. Self-assessment measures tend to be developed ad hoc and there are very few standardized measures with established reliability or validity. The face validity of this method is generally strong. The data are variable on the degree to which self-assessments correlate with ratings by peers or supervisors and with measures of performance (Dunning, Heath, & Suls, 2004; Eva, Cunnington, Reiter, Keane, & Norman, 2004; Fletcher & Baldry, 2000; Mattheos, Nattestad, Falk-Nilsson, & Attstrom, 2004; Swick et al., 2006). The least accurate self-assessments are

those of professionals who are least skilled and/or most confident (Davis et al., 2006).

Strengths. Self-assessment is associated with increases in self-knowledge on one's level of competency attainment across all foundational and functional competency domains. It has the potential to offer an effective approach for correcting self-appraisals when used in conjunction with ratings from other informants. If people are appropriately trained, the use of this method promotes self-reflection and self-reflective practice as a lifelong competency (Fletcher & Bailey, 2003).

Challenges. Self-assessment requires accurate self-appraisal, which is difficult to teach and to accomplish. Points of reference from which to conduct self-assessments often are lacking. It raises questions about the accuracy of the reporting, particularly as people with low rated competence according to other informants often overrate themselves. It is optimal when used in conjunction with other assessments.

Simulations/Role Plays

Description. Simulations/role plays involve constructing situations that resemble actual practice scenarios associated with various functional competencies, which are directly observed and evaluated with attention to foundational and functional competencies (Andrews, Brewin, & Rose, 2003; Bandiera et al., 2006; Manring et al., 2003). Simulations are the artificial recreation of a clinical environment or circumstance for the purpose of allowing the person being assessed to undertake a specific task, problem-solve, and receive immediate feedback to correct errors in a controlled fashion without endangering clients/patients. Simulations may incorporate the use of technology (e.g., virtual simulations or environments). Role plays involve taking on a role (e.g., as client/patient, or therapist) or putting oneself in another position for purposes of learning a new skill. Role plays typically include at least two people, such as the assessor/therapist/consultant/supervisor/teacher/manager/advocate and client/patient/student/consumer. Some authors have noted the utility of this approach for assessing a limited number of overall or broad and essential components of foundational and functional competencies (Bandiera et al., 2006; Epstein, 2007; Manring et al., 2003; Sidhu et al., 2004; Steadman et al., 2006; Swick et al., 2006).

Implementation. The competencies to be evaluated must first be identified. Then simulations for target competency(ies) should be constructed and the format (computer-based, standardized client/patient) of the simulations should be determined. Role plays for target competency(ies) should be created, role play partners selected, and the length/nature/complexity of role play scenario determined. The assessment method to be used must be determined and should include those already developed for use with role plays (e.g., microtraining skills). It is essential that the feedback mechanism to the person being assessed be established.

Psychometrics. Studies of the psychometric properties of simulations/role plays underscore its predictive validity. They demonstrate that the validity of such assessments is greater the closer the behavior is to the target competency and the more authentic the situation. Research reveals that construct validity is strengthened when the skill domain is assessed comprehensively. Furthermore, high quality simulations that closely resemble a real client/patient have good content validity. Interrater reliability increases with the

specificity of the coding scheme and degree of assessor training (Schuwirth & van der Vleuten, 2003; Swick et al., 2006). Case specificity poses the greatest threat to reliability (Schuwirth & van der Vleuten, 2003).

Strengths. This approach targets focally specific competencies for assessment. It provides useful structure for students or trainees to learn and practice new skills. It can be used reliably to assess competence in multiple domains (Scalese, Obeso, & Issenberg, 2008).

Challenges. This assessment method requires students or trainees to be informed prior to admission to a program that personal information may be expected to be disclosed during a role play. It raises dual role issues. This approach fails to capture how individuals actually perform in practice given the artificial nature of the situation and poses challenges for the individuals playing the client/patient role in terms of projecting genuine emotions or taking the role seriously. It creates challenges for people learning new skills if the individuals role playing clients/patients have more serious issues than novices can handle. Although role plays require limited resources or demands for implementation, simulations may be costly to create and implement (Manning et al., 2003).

Standardized Client/Patient Interviews

Description. With standardized client/patient interviews, individuals are trained to present as a patient/client with a specific condition in a standardized and consistent manner (Tamblyn, Klass, Schnabl, & Kopelow, 1991). The person being assessed then performs prescribed tasks such as interviewing, assessing, or intervening as though interacting with an actual client/patient. Assessors rate targeted competencies of persons being assessed during the conduct of a standardized client/patient interview using performance criteria set in advance. The use of standardized client/patient interviews may be a component of the aforementioned OSCE process or may occur in isolation, typically with one standardized client/patient interview that is longer in duration. Some have noted the usefulness of this methodology for summative evaluations (Epstein, 2007).

Implementation. To utilize standardized client/patient interviews, the target competency(ies) to be assessed must be determined. The measure(s) to be used must be ascertained. Checklists or rating forms need to be selected. Criteria setting needs to occur. The standardized client/patients must be trained. Directions need to be provided to the person being assessed from the assessor or from the standardized client/patient in order to conduct an evaluation of the standardized client/patient in a manner consistent with how they would act with an actual client/patient. Performance ratings must be secured in accord with the competencies that have been identified a priori from trained observers, the standardized client/patient, and the person being assessed.

Psychometrics. Using standardized clients/patients provides high fidelity assessments with reliable scores and good construct and concurrent validity. Reproducible scores are easy to obtain for the essential components of functional competencies. However, a number of cases/encounters need to be utilized to ensure reliability. Training of the standardized client/patient and the evaluators is key to reliability and validity, as they must be knowledgeable about the condition, understand the importance of consistent responses, and anticipate common questions and interactions

(Amano et al., 2004; Epstein & Hundert, 2002; Whelan et al., 2005).

Strengths. This assessment strategy offers a realistic and high fidelity assessment of actual performance. It allows the flexibility to have the standardized client/patient present in way that is most relevant to the competency being measured. A standardized client/patient interview allows for direct comparison across people being assessed (Howley, 2004; Kak et al., 2001; Petrusa, 2004; Zabar et al., 2004).

Challenges. This methodology requires considerable time, cost, and resources in design and implementation. It demands extensive training of raters to obtain reliable scores. This approach raises questions about the comparability of the behavior of the standardized client/patient versus what actually happens in practice settings, such as ongoing interventions. Further, it poses difficulties with regard to developing scoring systems that accurately capture performance (Epstein & Hundert, 2002; Howley, 2004; Kak et al., 2001; Manning et al., 2003).

Structured Oral Examinations

Description. Structured oral examinations are a method of performance evaluation in which an assessor or panel of assessors poses a series of questions orally in a systematic fashion to the person being assessed who is expected to respond orally and these responses are evaluated in accord with a priori established criteria (Andrews & Burruss, 2004; Bandiera et al., 2006). They may involve general questions addressing knowledge, skills, and attitudes and their integration associated with foundational and functional competencies; inquiry regarding material written previously by the person being assessed; presentation and discussion of standardized vignettes; discussion and analysis of live or recorded performances; and role playing, and so forth. The literature supports the effectiveness of such examinations for investigating, for summative evaluation purposes, the essential elements of a number of specific foundational and functional competencies and a limited number of overall or broad foundational competency domains (Epstein, 2007; Sidhu et al., 2004; Swick et al., 2006).

Implementation. Implementing structured oral examinations necessitates determining and communicating to the person being assessed the components of the oral examination and what needs to be provided by the person being assessed (e.g., written documents, recorded materials). It requires developing materials (e.g., probes, vignettes) to be used, if applicable. Standardizing administration, rating, and scoring procedures is key, as is the training of assessors to be consistent in their evaluations.

Psychometrics. This approach has acceptable reliability and validity (face, content, construct, criterion), which is stronger the more standardized and structured the protocol (Sidhu et al., 2004; Swick et al., 2006). However, it is worth noting that oral examinations for purposes of licensure have been challenged in courts and some jurisdictions have eliminated the procedure. In addition, vignettes may have limited predictive validity (Epstein, 2007; Landon et al., 2003).

Strengths. The approach enables the assessor(s) to explore competencies in an in-depth manner. It offers the ability to assess the integration of knowledge, skills, and attitudes. It supports the assessment of the application of knowledge in decision-making. It also allows for a dynamic interaction between the assessor(s) and

the person being assessed. It enhances the reliability of the oral examination process beyond that found with unstructured oral examinations (Bandiera et al., 2006).

Challenges. This method requires considerable time on the part of all parties involved. It necessitates the training of assessors to systematically administer and rate the examination and to avoid giving verbal or nonverbal cues to the person being assessed. Standardizing administration both within and across assessors is critical. This approach poses challenges for assessing quality of care over time (Manring et al., 2003). In addition, there are questions about fidelity (Epstein, 2007; Landon et al., 2003).

Written Examinations

Description. Paper- or computer-based written examinations may include multiple choice, matching, fill-in-the-blank, true false questions; vignettes or problem-based scenarios requiring sequential and integrative problem-solving and decision-making; and essay questions that require the synthesis and communication of content and critical thinking and judgment (Bandiera et al., 2006). An extensively used summative evaluation tool, they can evaluate the essential elements of a number of foundational and functional competencies (individual and cultural diversity, interdisciplinary systems, assessment, intervention, consultation, research and evaluation, teaching; Epstein, 2007; Sidhu et al., 2004). They provide a reasonable way to tap the overall or broad foundational competencies in a limited number of domains (scientific knowledge and methods, ethical and legal standards and policy). Written examinations are most useful readiness to practicum to entry to practice developmental levels, but in some instances (e.g., board certification in certain specialties) may be used for advanced credentialing.

Implementation. The implementation of the written examination process consists of determining the competencies or essential subcomponents to be assessed and developing the questions and statements that the person being assessed will respond to. The exam is then administered and scored. The cutoff scores or grading curve are determined. The outcome (and if appropriate the test answers) of the written examination must be provided to the person being assessed.

Psychometrics. Given the lack of consistency of written examinations among graduate programs, no overall statement of psychometric properties is appropriate. However, psychometric examinations of the Examination of Professional Practice in Psychology (EPPP) reveal good content validity and that the various characteristics of the person being assessed's doctoral program correlate with EPPP scores (Greenberg & Jesuitus, 2003; Leigh et al., 2007). In addition, performance on written examinations correlates with performance on more observable assessment methods, such as OSCEs (Ram et al., 1999).

Strengths. This approach provides an easy to use and inexpensive method of competency assessment. It allows for the assessment of multiple people at the same time. This method offers an effective assessment of knowledge across multiple competency domains. It allows for easy and generally reliable scoring, and permit easy interpretation (Kak et al., 2001; Sidhu et al., 2004; Wass et al., 2001).

Challenges. Such an examination poses challenges for assessing skills and attitudes. It raises questions about applicability to real world experiences and thus the fidelity of the method. It yields

concerns about the predictive validity of performance on written examinations to performance in professional practice settings (Epstein, 2007; Sidhu et al., 2004; Swick et al., 2006).

Discussion

Recent years have witnessed a burgeoning of interest in identifying the foundational and functional competencies in professional psychology (Kaslow, 2004; Kaslow et al., 2004; Rodolfa et al., 2005) and in articulating benchmarks for competency acquisition at each stage of the education, training, and credentialing sequence (Fouad et al., 2009). In keeping with these efforts, the assessment of competence is gaining widespread attention within the profession (Kaslow, Rubin, Bebeau, et al., 2007; Nelson, 2007; Peterson, 2004; Roberts et al., 2005). This article offers tools that psychologists can include in their competency assessment armamentarium. A number of factors should guide decisions regarding appropriate assessment methods (Kak et al., 2001; Kaslow, Celano, & Stanton, 2005; Kaslow, Dunn, & Smith, 2008; Sidhu et al., 2004). Choices regarding which assessment methods to utilize should be based on the competency(ies) to be assessed. They should take into consideration the availability of psychometrically sound methods. The feasibility, strengths, and challenges of the methodology should be taken into account when deciding upon which assessment strategies should be utilized. Attention also should be paid to the appropriateness of the coding, scoring, and interpretation of the data gleaned. Furthermore, consideration should be given to the person(s) who will serve as the evaluator(s). In addition, the context of the assessment should guide decision-making. Although it is necessary to consider the psychometric properties of the approach for all evaluations, reliability, validity, and fidelity of the measurement is particularly key for high stakes assessments, which typically occur in the form of summative evaluations (Hodges et al., 2002).

No single strategy adequately assesses the overall and broad or essential components of all competencies. Each approach has its own strengths and challenges. Thus, the comprehensive assessment of the broad range of foundational and functional competencies and their essential components should incorporate multiple tools rated by multiple informants longitudinally (Sidhu et al., 2004). Using a combination of assessment tools will enhance the validity of the assessment process (Brown & Doshi, 2006). Furthermore, these multidimensional broader assessments should be conducted in accord with the guiding principles for the assessment of competence (Kaslow, Rubin, Bebeau, et al., 2007).

We recommend the following future directions for moving forward the assessment of competence toolkit. First, it behooves us at a national level to develop standardized tools, protocols, and procedures that can be used for different stages of professional development (Schuwirth & van der Vleuten, 2003). Such assessment measures must tap the relevant foundational and functional competencies within professional psychology and do so in a fashion that can be used systematically across training programs and credentialing efforts. The methods developed must be reasonably feasible and cost-efficient. These templates could build on already existing models within professional psychology, such as the Self-Assessment Guide and Professional Development Plan completed biannually by psychologists in Ontario (<http://www.cpo.on.ca/>), or from protocols that are used in other health professions (e.g.,

OSCEs; Gaba, 2004). For some common assessment methods, it is advisable that scoring procedures be designed and shared across programs. For certain tools, it would be advantageous if centralized systems for scoring and writing evaluation reports be developed (e.g., 360-degree evaluations), as this could help offset some of the costs of implementation. More psychometric analyses need to be performed to ascertain the reliability, validity, and fidelity of the various assessment methods and specific measurement tools within professional psychology. More attention should be paid to adapting assessment methods that traditionally are used for formative evaluation to summative evaluation and vice versa. This will become increasingly possible as methods are crafted and shared and more formal methods for rating/scoring assessments become available. These tools, protocols, and procedures should be shared across education and training programs and credentialing bodies and disseminated widely.

Second, guidelines for selecting among the various set of measurement approaches is an area worthy of future exploration. The development of such guidelines would require attention to the strengths and challenges associated with each assessment approach, the psychometric soundness of specific measurement tools that are developed in keeping with each overarching assessment strategy, the stage of professional development for which the assessment battery is being utilized, the context(s) of the assessment, and the competencies being assessed.

Third, new technologies also can be used to support the advancement of innovative approaches. For example, some of the assessment methods that appear in the toolkit could be updated and enhanced through the use of technology. Portfolios could include EPortfolios, which are easier for the person being assessed to use (Wickersham & Chambers, 2006). These EPortfolios also may prove quite valuable for documenting continuing professional development beyond licensure. Furthermore, simulator technologies are very promising for teaching competencies and assessing these competencies via the use of reproducible stimuli (Srinivasan, Hwang, West, & Yellowlees, 2006). One simulator technology that has gained attention is the use of virtual clients/patients, an approach that may be used for both teaching foundational and functional competencies and the assessment of these competencies (Srinivasan et al., 2006; Stevens et al., 2006). This approach may offer trainees a safe, secure, consistent learning and evaluation context (Stevens et al., 2006). Virtual reality simulation also may prove to be a very valuable strategy for performance assessment, and in other health professions this approach has been found to have good psychometric properties (Patel, Gallagher, Nicholson, & Cates, 2006). There also are simulation technologies, such as Strategic Management Simulation, that can be borrowed from professions other than health care professions that could bolster our use of technology to assess competence (Satish, Streufert, & Barach, 2001). Increasingly, technology should be considered our friend as we advance and enhance our assessment armamentarium.

Fourth, and in a related vein, relevant theories and empirical research may serve as valuable guiding frameworks for the creation of effective assessment techniques. The following are two such examples. Instructional design, which incorporates principles of human learning can be utilized not only for the purposes of designing instructional systems, but also for authentic performance assessment that is linked to the mastery criteria (i.e., benchmarks) associated with each learning outcome (Dick, Carey, & Carey,

2004). Gaming theory offers a wealth of resources and tools that may be adapted and modified to guide the creation of more sophisticated and relevant simulations/role plays (Satish et al., 2001; Srinivasan et al., 2006).

Fifth, training modules and/or programs need to be created for each assessment approach to ensure that assessors appropriately, validly, and reliably rate the competence of the people they are assessing, as well as effectively provide competency assessment feedback (Brown & Doshi, 2006). Similarly, more systematic attention needs to be paid to devising and disseminating strategies for training students and professionals to accurately and meaningfully engage in the process of self-assessment (Belar et al., 2001; Loacker, 2000; Roberts et al., 2005; Stewart et al., 2000; Swick et al., 2006).

Sixth, there is a need to work together to develop a common approach to documenting the maintenance of competence (Brown, Miles, Perelman, & Stockman, 2009). As more attention is paid to the recertification of professional psychologists, consideration may be given to the use of a wide array of approaches that will be included as a component of the portfolio that would be used for such recertification efforts (Norcini, 1999).

Developing and implementing assessment methods to evaluate competence remains a challenge for psychology educators, trainers, regulators, and credentialing bodies. We need to adapt existing tools and create new tools to meet this challenge (Murray, Gruppen, Catton, Hays, & Wooliscroft, 2000). Hopefully, this Competency Assessment Toolkit for Professional Psychology will support the inclusion of a broad array of assessment tools, rated by multiple informants (e.g., self, supervisors, peers, subordinates, consumers, etc.) for formative and summative assessment throughout the professional life cycle and across practice settings. To date, most of the literature on specific strategies has been provided by other health professionals, and the time is now for professional psychologists to develop, evaluate, and disseminate best practice methods for assessing the core foundational and functional competencies. Professional psychologists need to ascertain the fidelity of the pedagogical paradigm shift in education and training from a focus on the objectives of the learning process to outcomes based learning and competency-based assessments in terms of producing more competent professional psychologists (Albanese, Mejicano, Mullan, Kokotailo, & Gruppen, 2008; Carraccio, Wolfsthal, Englander, Ferentz, & Martin, 2002). Furthermore, creating and implementing more comprehensive and effective evaluation approaches needs to involve the regulatory and credentialing communities to enhance the assessment of competency throughout the sequence of training leading to licensure.

References

- Accreditation Council for Graduate Medical Education and American Board of Medical Specialties. (2000). *Toolbox of assessment methods: A product of the joint initiative of the ACGME Outcome Project of the Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS), version 1.1*. Chicago: Accreditation Council for Graduate Medical Education and American Board of Medical Specialties.
- Adamo, G. (2003). Simulated and standardized patients in OSCEs: Achievements and challenges 1992-2003. *Medical Teacher*, 25, 262-270.
- Albanese, M. A., Mejicano, G., Mullan, P., Kokotailo, P., & Gruppen, L.

- (2008). Defining characteristics of educational competencies. *Medical Education*, 42, 248–255.
- Amano, H., Sano, T., Gotoj, K., Kahuta, S., Suganuma, T., Kimura, Y., et al. (2004). Strategies for training standardized patient instructors for a competency exam. *Journal of Dental Education*, 68, 1104–1111.
- Andrews, B., Brewin, C. R., & Rose, S. (2003). Gender, social support, and PTSD in victims of violent crime. *Journal of Traumatic Stress*, 16, 421–427.
- Andrews, L. B., & Burruss, J. W. (2004). *Core competencies for psychiatric education: Defining, teaching, and assessing resident competence*. Washington, DC: American Psychiatric Publishing Inc.
- Arredondo, P., Shealy, C., Neale, M. C., & Winfrey, L. L. (2004). Consultation and interprofessional collaboration: Modeling for the future. *Journal of Clinical Psychology*, 80, 787–800.
- Atkins, P. W. B., & Wood, R. E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology*, 55, 871–904.
- Bandiera, G., Sherbino, J., & Frank, J. R. (Eds.). (2006). *The CanMEDS assessment tools handbook: An introductory guide to assessment methods for the CanMEDS competencies*. Ottawa, Ontario: The Royal College of Physicians and Surgeons of Canada.
- Belar, C. D., Brown, R. A., Hersch, L. E., Hornyak, L. M., Rozensky, R. H., Sheridan, E. P., et al. (2001). Self-assessment in clinical health psychology: A model for ethical expansion of practice. *Professional Psychology: Research and Practice*, 32, 135–141.
- Bienenfeld, D., Klyklyo, W., & Knapp, V. (2000). Process and product: Development of competency-based measures for psychiatry residency. *Academic Psychiatry*, 24, 68–76.
- Bieschke, K. J., Fouad, N. A., Collins, F. L., & Halonen, J. S. (2004). The scientifically-minded psychologist: Science as a core competency. *Journal of Clinical Psychology*, 80, 713–724.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–75.
- Brown, H. J., Miles, P. V., Perelman, R. H., & Stockman, J. A. (2009). A continuum of competency assessment: The potential for reciprocal use of the Accreditation Council for Graduate Medical Education Toolbox and the components of the American Board of Pediatrics Maintenance-of-Certification Program. *Pediatrics*, 123, 556–558.
- Brown, N., & Doshi, M. (2006). Assessing professional and clinical competence: The way forward. *Advances in Psychiatric Treatment*, 12, 81–91.
- Carraccio, C., & Englander, R. (2000). The Objective Structured Clinical Examination: A step in the direction of competency-based evaluation. *Archives of Pediatrics and Adolescent Medicine*, 154, 736–741.
- Carraccio, C., & Englander, R. (2004). Evaluating competence using a portfolio: A literature review and web-based application to the ACGME competencies. *Teaching and Learning in Medicine*, 16, 381–387.
- Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: From Flexner to competencies. *Academic Medicine*, 77, 361–367.
- Carson, M. (2006). Saying it like it isn't: The pros and cons of 360-degree feedback. *Business Horizons*, 49, 395–402.
- Caverzagie, K. J., Shea, J. A., & Kogan, J. R. (2008). Resident identification of learning objectives after performing self-assessment based upon the ACGME core competencies. *Journal of General Internal Medicine*, 23, 1024–1027.
- Chin, J. L. (2000). Culturally competent health care. *Public Health Reports*, 115, 25–33.
- Colliver, J. A., Willis, M. S., Robbs, R. S., Cohen, D. S., & Swartz, M. H. (1998). Assessment of empathy in a standardized-patient examination. *Teaching and Learning in Medicine*, 10, 8–11.
- Cowan, D. T., Norman, I., & Coopamah, V. P. (2005). Competence in nursing practice: A controversial concept—A focused review of literature. *Nurse Education Today*, 25, 355–362.
- Daniel, J. H., Roysircar, G., Abeles, N., & Boyd, C. (2004). Individual and cultural diversity competency: Focus on the therapist. *Journal of Clinical Psychology*, 80, 755–770.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Harrison, R. V., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association*, 296, 1094–1102.
- de las Fuentes, C., Willmuth, M. E., & Yarrow, C. (2005). Ethics education: The development of competence, past and present. *Professional Psychology: Research and Practice*, 36, 362–366.
- Dick, W., Carey, L., & Carey, J. (2004). *Systematic design of instruction*. Upper Saddle River, NJ: Pearson Education.
- Dingle, A. D., & Beresin, E. (2006). Competencies. *Child and Adolescent Psychiatric Clinics of North America*, 16, 225–247.
- Driessen, E., Van der Vleuten, C., Schuwirth, L., van Tartwijk, J., & Vermunt, J. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Medical Education*, 39, 214–220.
- Dufrene, R. L. (2000). An evaluation of a patient satisfaction survey: Validity and reliability. *Evaluation and Program Planning*, 23, 293–300.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5, 69–106.
- Elman, N., Illfelder-Kaye, J., & Robiner, W. (2005). Professional development: A foundation for psychologist competence. *Professional Psychology: Research and Practice*, 36, 367–375.
- Epstein, R. M. (2007). Assessment in medical education. *New England Journal of Medicine*, 356, 387–396.
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, 287, 226–235.
- Eva, K. W., Cunnington, J. P. W., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Sciences Education*, 9, 211–224.
- Evans, R., Elwyn, G., & Edwards, A. (2004). Review of instruments for peer assessment of physicians. *British Medical Journal*, 328, 1240–1243.
- Evans, R. G., Edwards, A., Evans, S. A., Elwyn, B., & Elwyn, G. (2007). Assessing the practising physician using patient surveys: A systematic review of instruments and feedback methods. *Family Practice*, 24, 117–127.
- Falchikov, N., & Goldfinch, J. (2000). Student-peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.
- Falender, C. A., Cornish, J. A. E., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., et al. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology*, 80, 771–786.
- Fernsten, L., & Fernsten, J. (2005). Portfolio assessment and reflection: Enhancing learning through effective practice. *Reflective Practice*, 6, 303–309.
- Fletcher, C. E., & Bailey, C. (2003). Assessing self-awareness: Some issues and methods. *Journal of Managerial Psychology*, 18, 395–404.
- Fletcher, C. E., & Baldry, C. (2000). A study of individual differences and self-awareness in the context of multi-source feedback. *Journal of Occupational and Organizational Psychology*, 73, 303–319.
- Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M., et al. (2009). Competency benchmarks: A model for the understanding and measuring of competence in professional psychology across training levels. *Training and Education in Professional Psychology*, 4(Suppl.), S5–S26.
- Fryer-Edwards, K., Pinsky, L. E., & Robins, L. (2006). The use of port-

- folios to assess professionalism. In D. T. Stern (Ed.), *Measuring medical professionalism* (pp. 213–233). Oxford: Oxford University Press.
- Gaba, D. M. (2004). The future of simulation in health care. *Quality and Safety of Health Care, 13*(Suppl), 2–10.
- Greenberg, S., & Jesuitis, L. (2003). *Study of the practice of licensed psychologists in the United States and Canada prepared for the Association of State and Provincial Psychology Boards*. New York: Professional Examination Service Department of Research and Development.
- Grol, R. (2001). Improving the quality of medical care: Building bridges among professional pride, payer profit, and patient satisfaction. *Journal of the American Medical Association, 286*, 2578–2585.
- Hatcher, R. L., & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*, 12–25.
- Hatcher, R. L., & Lassiter, K. D. (2007). Initial training in professional psychology: The practicum competencies outline. *Training and Education in Professional Psychology, 1*, 49–63.
- Hobgood, C. D., Riviello, R. J., Jouriles, N., & Hamilton, G. (2002). Assessment of communication and interpersonal skills competencies. *Academic Emergency Medicine, 9*, 1257–1269.
- Hodges, B., Hanson, M., McNaughton, N., & Regehr, G. (2002). Creating, monitoring, and improving a psychiatry OSCE. *Academic Psychiatry, 26*, 134–161.
- Hoge, M. A., Morris, J. A., Daniels, A. S., Huey, L. Y., Stuart, G. W., Adams, N., et al. (2005). Report of recommendations: The Annapolis Coalition on Behavioral Health Work Force Competencies. *Administration and Policy in Mental Health, 32*, 651–663.
- Holmboe, E. S., Hawkins, R. E., & Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence: A randomized trial. *Annals of Internal Medicine, 140*, 874–881.
- Horvath, A. O., & Greenberg, L. S. (1986). The development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529–556). New York: Guilford Press.
- Howley, L. D. (2004). Performance assessment in medical education: Where we've been and where we're going. *Evaluation & The Health Professions, 27*, 285–303.
- Joshi, R., Ling, F. W., & Jaeger, J. (2004). Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Academic Medicine, 79*, 458–463.
- Jouriles, N. J., Burdick, W., & Hobgood, C. D. (2002). Clinical assessment in emergency medicine. *Academic Emergency Medicine, 9*, 1289–1294.
- Jouriles, N. J., Emerman, C. L., & Cydulka, R. K. (2002). Direct observation for assessing emergency medicine core competencies: Interpersonal skills. *Academic Emergency Medicine, 9*, 1338–1341.
- Kak, N., Burkhalter, B., & Cooper, M.-A. (2001). *Measuring the competence of healthcare providers*. Bethesda, MD: Published for the U.S. Agency for International Development (USAID) by the Quality Assurance (QA) Project.
- Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist, 59*, 774–781.
- Kaslow, N. J., & Bell, K. D. (2008). A competency-based approach to supervision. In C. A. Falender & E. P. Shafranske (Eds.), *Casebook for clinical supervision: A competency-based approach* (pp. 17–38). Washington, DC: American Psychological Association.
- Kaslow, N. J., Borden, K. A., Collins, F. L., Forrest, L., Illfelder-Kaye, J., Nelson, P. D., et al. (2004). Competencies Conference: Future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology, 80*, 699–712.
- Kaslow, N. J., Celano, M. P., & Stanton, M. (2005). Training in family psychology: A competencies-based approach. *Family Process, 44*, 337–353.
- Kaslow, N. J., Dunn, S. E., & Smith, C. O. (2008). Competencies for psychologists in academic health centers (AHCs). *Journal of Clinical Psychology in Medical Settings, 15*, 18–27.
- Kaslow, N. J., Rubin, N. J., Bebeau, M., Leigh, I. W., Lichtenberg, J., Nelson, P. D., et al. (2007). Guiding principles and recommendations for the assessment of competence. *Professional Psychology: Research and Practice, 38*, 441–451.
- Kaslow, N. J., Rubin, N. J., Forrest, L., Elman, N. S., Van Horne, B. A., Jacobs, S. C., et al. (2007). Recognizing, assessing, and intervening with problems of professional competence. *Professional Psychology: Research and Practice, 38*, 479–492.
- Krishnamurthy, R., Vandecreek, L., Kaslow, N. J., Tazeau, Y. N., Milville, M. L., Kerns, R., et al. (2004). Achieving competency in psychological assessment: Directions for education and training. *Journal of Clinical Psychology, 80*, 725–740.
- Landon, B. E., Normand, S.-L. T., Blumenthal, D., & Daley, J. (2003). Physician clinical performance assessment: Prospects and barriers. *Journal of the American Medical Association, 290*, 1183–1189.
- Leigh, I. W., Smith, I. L., Bebeau, M., Lichtenberg, J., Nelson, P. D., Portnoy, S., et al. (2007). Competency assessment models. *Professional Psychology: Research and Practice, 38*, 463–473.
- Lichtenberg, J., Portnoy, S., Bebeau, M., Leigh, I. W., Nelson, P. D., Rubin, N. J., et al. (2007). Challenges to the assessment of competence and competencies. *Professional Psychology: Research and Practice, 38*, 474–478.
- Lievens, F., & Sanchez, J. I. (2007). Can training improve the quality of inferences made by raters in competency modeling? A quasi-experiment. *Journal of Applied Psychology, 92*, 812–819.
- Loacker, G. (Ed.). (2000). *Self assessment at Alverno College*. Milwaukee, WI: Alverno College.
- Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *Journal of Continuing Education in the Health Professions, 23*, 4–12.
- Lynch, D. C., Swing, S. R., Horowitz, S. D., Holt, K., & Messer, J. V. (2004). Assessing practice-based learning and improvement. *Teaching and Learning in Medicine, 16*, 85–92.
- Manning, J., Beitman, B. D., & Dewan, M. J. (2003). Evaluating competence in psychotherapy. *Academic Psychiatry, 27*, 136–144.
- Mariush, M. E. (Ed.). (2004). *The use of psychological testing for treatment planning and outcomes assessment* (3rd ed.). Mahwah, NJ: Erlbaum.
- Mattheos, N., Nattestad, A., Falk-Nilsson, E., & Attström, R. (2004). The interactive examination: Assessing students' self-assessment ability. *Medical Education, 38*, 378–389.
- Medical School Objectives Writing Group. (1999). Learning objectives for medical student education—Guidelines for medical schools: Report I of the Medical School Objectives Project. *Academic Medicine, 74*, 13–18.
- Murray, E., Gruppen, L., Catton, P., Hays, R. B., & Wooliscroft, J. O. (2000). The accountability of clinical education: Its definition and assessment. *Medical Education, 34*, 871–879.
- Nelson, P. D. (2007). Striving for competence in the assessment of competence: Psychology's professional education and credentialing journey of public accountability. *Training and Education in Professional Psychology, 1*, 3–12.
- Newble, D. (2004). Techniques for measuring clinical competence: Objective structured clinical examinations. *Medical Education, 38*, 199–203.
- Norcini, J. J. (1999). Recertification in the United States. *British Medical Journal, 319*, 1183–1185.
- Patel, A. D., Gallagher, A. G., Nicholson, W. J., & Cates, C. U. (2006). Learning curves and reliability measures for virtual reality simulation in the performance assessment of carotid angiography. *Journal of the American College of Cardiology, 47*, 1796–1802.
- Peterson, R. L. (2004). Evaluation and the cultures of professional psychology education programs. *Professional Psychology: Research and Practice, 35*, 420–426.

- Peterson, R. L., Peterson, D. R., Abrams, J. C., & Stricker, G. (1997). The National Council of Schools and Programs of Professional Psychology education model. *Professional Psychology: Research and Practice, 28*, 373-386.
- Petrusa, E. R. (2004). Taking standardized patient-based examinations to the next level. *Teaching and Learning in Medicine, 16*, 98-110.
- Petti, P. (2008). The use of a structured case presentation examination to evaluate clinical competencies of psychology doctoral students. *Training and Education in Professional Psychology, 2*, 145-150.
- Pitts, J., Coles, C., & Thomas, P. (1999). Educational portfolios in the assessment of general practice trainers: Reliability of assessors. *Medical Education, 33*, 515-520.
- Pope, K. S., Sonne, J. L., & Greene, B. (2006). *What therapists don't talk about and why: Understanding taboos that hurt us and our clients*. Washington, DC: American Psychological Association.
- Ram, P., Grol, R., Rethans, J. J., Schouten, B., van der Vleuten, C., & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: Issues of validity, reliability, and feasibility. *Medical Education, 33*, 447-454.
- Roberts, M. C., Borden, K. A., Christiansen, M. D., & Lopez, S. J. (2005). Fostering a culture shift: Assessment of competence in the education and careers of professional psychologists. *Professional Psychology: Research and Practice, 36*, 355-361.
- Rodgers, K. G., & Manifold, C. (2002). 360-degree feedback: Possibilities for assessment of the ACGME core competencies for emergency medicine residents. *Academic Emergency Medicine, 9*, 1300-1304.
- Rodolfa, E. R., Bent, R. J., Eisman, E., Nelson, P. D., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice, 36*, 347-354.
- Rubin, N. J., Bebeau, M., Leigh, I. W., Lichtenberg, J., Smith, I. L., Nelson, P. D., et al. (2007). The competency movement within psychology: An historical perspective. *Professional Psychology: Research and Practice, 38*, 452-462.
- Satish, U., Streufert, S., & Barach, P. (2001). Assessing and improving medical competency: Using strategic management simulations. *Simulation & Gaming, 32*, 156-163.
- Scalese, R. J., Obeso, V. T., & Issenberg, B. (2008). Simulation technology for skills training and competency assessment in medical education. *Journal of General Internal Medicine, 23*(Suppl), 46-49.
- Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2003). The use of clinical simulations in assessment. *Medical Education, 37*(Suppl 1), 65-71.
- Sidhu, R. S., Grober, E. D., Musselman, L. J., & Reznick, R. K. (2004). Assessing competency in surgery: Where to begin? *Surgery, 135*, 6-20.
- Spielman, A. I., Fulmer, T., Eisenberg, E. S., & Alfano, M. C. (2005). Dentistry, nursing, and medicine: A comparison of core competencies. *Journal of Dental Education, 69*, 1257-1271.
- Spruill, J., Rozensky, R. H., Stigall, T. T., Vasquez, M., Bingham, R. P., & Olvey, C. D. V. (2004). Becoming a competent clinician: Basic competencies in intervention. *Journal of Clinical Psychology, 80*, 741-754.
- Srinivasan, M., Hwang, J. C., West, D., & Yellowlees, P. M. (2006). Assessment of clinical skills using simulator technologies. *Academic Psychiatry, 30*, 505-515.
- Steadman, R. H., Coates, W. C., Huang, Y. M., Matevosian, R., Larmon, B. R., McCullough, L., et al. (2006). Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Critical Care Medicine, 34*, 151-157.
- Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Rajj, A., Harrison, C., et al. (2006). The use of virtual patients to teach medical students history taking and communication skills. *The American Journal of Surgery, 191*, 806-811.
- Stewart, J., O'Halloran, C., Barton, J. R., Singleton, S. J., Harrigan, P., & Spencer, J. (2000). Clarifying the concepts of confidence and competence to produce appropriate self-evaluation measurement scales. *Medical Education, 34*, 903-909.
- Swick, S., Hall, S., & Beresin, E. (2006). Assessing the ACGME competencies in psychiatry training programs. *Academic Psychiatry, 30*, 330-351.
- Tamblyn, R. M., Klass, D. J., Schnabl, G. K., & Kopelow, M. L. (1991). The accuracy of standardized patient presentation. *Medical Education, 25*, 100-109.
- Tate, P., Foulkes, J., Neighbour, R., Champion, P., & Field, S. (1999). Assessing physicians' interpersonal skills via videotaped encounters: A new approach for the Royal College of General Practitioners Membership examination. *Journal of Health Communication, 4*, 143-152.
- Townsend, A. H., McIlvenny, S., Miller, C. J., & Dunn, E. V. (2001). The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Medical Education, 35*, 841-846.
- Wass, V., van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Medical Education Quartet, 357*, 945-949.
- Watson, R., Stimpson, A., Topping, A., & Porock, D. (2002). Clinical competence assessment in nursing: A systematic review of the literature. *Advanced Nursing, 39*, 421-431.
- Weigelt, J. A., Brasel, K. J., Bragg, D., & Simpson, D. (2004). The 360-degree evaluation: Increased work with little return? *Current Surgery, 61*, 616-628.
- Whelan, G. P., Boulet, J. R., McKinley, D. W., Norcini, J. J., van Zanten, M., Hambleton, R. K., et al. (2005). Scoring standardized patient examinations: Lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA). *Medical Teacher, 27*, 200-206.
- Wickersham, L. E., & Chambers, S. M. (2006). EPortfolios: Using technology to enhance and assess student learning. *Education, 126*, 738-746.
- Wilkinson, T. J., Challis, M., Hobma, S. O., Newble, D. I., Parboosingh, J. T., Sibbald, R. G., et al. (2002). The use of portfolios for assessment of the competence and performance of doctors in practice. *Medical Education, 36*, 918-924.
- Zabar, S., Hanley, K., Stevens, D. L., Kalet, A., Schwartz, M. D., Pearlman, E., et al. (2004). Measuring the competence of residents as teachers. *Journal of General Internal Medicine, 19*, 530-533.

Received September 2, 2008

Revision received February 10, 2009

Accepted February 17, 2009 ■