Assessing Psychologists in Practice: Lessons From the Health Professions Using Multisource Feedback

Jac J. W. Andrews, Claudio Violato, Ahmed Al Ansari, Tyrone Donnon, and Gia Pugliese
The University of Calgary

The aim of the present study was to conduct a systematic literature review of multisource feedback (MSF) instruments and to summarize the evidence of feasibility, reliability, generalizability, validity, and other psychometric characteristics of the instruments. Accordingly, we conducted a systematic literature review for English-language studies published from 1975 to 2012 using the following databases: MEDLINE, EMBASE, CINAHL, PubMed, and PsycINFO. The following terms were used in the search: multisource feedback, 360-degree evaluation, and assessment of medical professionalism. Forty-eight studies conducted in Canada, the United States, the United Kingdom, Netherlands, China, and elsewhere met the inclusion criteria. The results indicate that MSF has adequate evidence of validity, reliability, and feasibility for providing health practitioners with quality improvement data (both formative and summative assessment) as part of an overall strategy of maintaining competence and certification. Professional psychology has not adopted MSF as a systematic competence-based method for evaluating, maintaining, and assuring competent practice of psychology and instead relies on self-assessment as the primary quality assurance approach for its public accountability. We make recommendations to adopt an MSF system of competence-based assessment of practicing psychologists by regulatory and licensing authorities in Canada and the United States.

Keywords: multisource feedback, 360-degree evaluation, professional competence, psychologist assessment, formative assessment

A major professional issue for psychology is the monitoring and assessment of psychologists in professional practice. Multisource feedback (MSF), sometimes referred to as 360-degree assessment, has become increasingly common in health care settings. For example, it has been widely used in industry as a way of providing feedback to employees in order to improve workplace performance and guide self-directed learning (Sala & Dwight, 2002). However, the use of multiple sources of information to evaluate performance has also been considered useful in business (Sala & Dwight, 2002). For example, it has been used by supervisors in the form of providing feedback to employees to guide self-directed learning (Sala & Dwight, 2002) when supervisors and peers provide MSF. Often, MSF is used in settings in which the staff person works in a team and/or cannot be directly supervised (Church, 1997). Similarly, in psychology, multiple sources of input can be valuable when evaluating the same dimension such as the case of a trainee’s development as a therapist (Falender & Shafranske, 2004a). Cone (2001), for example, suggested the use of measurement of multiple aspects of a client’s constructs to judge their relative effectiveness. More generally, Falender and Shafranske (2004b) believe the use of MSF provides an interesting analysis of training performance within the context of clinical supervision and psychology training. Although MSF has not been used as a way to evaluate the performance of practicing psychologists, Andrews and Violato (2010) presented information about how MSF instruments could be constructed and used as a competence-monitoring system for practicing school psychologists and what these instruments could look like. In the present article, we propose the use of MSF for the assessment of psychologists in practice. We reviewed developments and research findings in health care and psychology to provide some considerations and directions for MSF in professional psychology.
Professional Competency and Assessment of Psychologists: An Overview

Psychologists are generally thought to be competent if they have the knowledge, skills, attitudes, and professionalism for practice (e.g., assessment, intervention, consultation; Sharpless & Barber, 2009). According to Epstein and Hundert (2002), competence is “the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served” and is dependent on “habits of mind, including attentiveness, critical curiosity, self-awareness, and presence” (p. 227). According to the Continuing Competence Program for Psychologists Practicing in Nova Scotia (http://www.nsbep.org/downloads/Continuing_Competence) in Canada, competence for psychologists involves the interaction of four major components: knowledge (of a range of professional issues), skill (ability to apply knowledge), judgment (when and where to apply skills), and diligence (consistent application of knowledge, skills, and judgment). Moreover, competence is the potential for appropriate professional practice distinct from performance in daily practice, which is situation specific, and should be observable, measurable, and developmental.

The conceptualization and measurement of performance and competence in the health professions have been evolving processes over the years. Performance and competence are fluid states whereby practice standards are developed throughout one’s professional education, training, and experience. Health professions (including psychology) have established foundations and criteria for determining professional competence including ethical principles and standards of practice (e.g., American Psychological Association, 2002; Canadian Psychological Association, 2000), registration and licensure requirements through professional credentialing bodies (e.g., Association of State and Provincial Psychology Boards, which is the alliance of state, provincial, and territorial agencies responsible for the licensure and certification as well as the maintenance of competence of psychologists throughout the United States and Canada; http://www.asppb.net), education and training (e.g., American Psychological Association- and Canadian Psychological Association-accredited programs), and specialty board certification processes and procedures (e.g., American Board of Clinical Psychology, American Board of Counseling Psychology, American Board of School Psychology).

Over the past several decades, psychology has moved toward considering, developing, and implementing valid, reliable, and feasible assessment of performance and competence in order for the profession to improve its public accountability and as a way to provide both formative and summative feedback to psychologists to improve, enrich, or confirm their performance consistent with expected outcomes (Carraccio, Wolfshlat, Englander, Ferenz, & Martin, 2002; Roberts, Borden, Christiansen, & Lopez, 2005). Professional competence in health care is expected by patients and clients, as well as policymakers and regulators (Hoge et al., 2005). Accordingly, the assessment of competence and performance protects the public and encourages professional development (Kaslow, 2004). Although there has been recent support for the systematic assessment of competence in professional psychology (Kaslow, Rubin, Bebeau, et al., 2007), there is still a lack of consensus in the field about how to define and measure the competence of professional psychologists in practice (Fantuzzo, Sisemore, & Spradlin, 1983; Kaslow, 2004; Shaw & Dobson, 1988). Moreover, licensing and regulatory boards as well as graduate training programs also vary in their competence-based assessment approach (Roberts et al., 2005). Although licensing and regulatory boards in Canada and the United States have not adopted agreed-on principles and objectives to assess the practice of psychologists, there have been recommendations made in psychology for the assessment of competence.

Kaslow et al. (2009), for example, presented a competence assessment toolkit for professional psychology in which various assessment instruments along with information regarding their implementation, reliability, validity, and fidelity were reported. They reviewed several methods deemed appropriate for assessing the foundational competencies (e.g., professionalism) and functional competencies (e.g., assessment) of psychologists and for measuring competence for education, training, and professional development. The methods described and reviewed included annual/rotation performance reviews, case presentation reviews, client/patient process and outcome data, competence evaluation rating forms, consumer surveys, live or recorded performance ratings, objective structured clinical examinations, portfolios, record reviews, self-assessment, simulations/role plays, standardized client/patient interviews, structured oral examinations, written examinations, and 360-degree evaluations.

In addition to providing descriptions and reviews of these assessment methods, Kaslow et al. (2009) noted the following considerations for choosing and using one or more of the methods: (a) types of competencies to be assessed; (b) psychometric properties of the methods; (c) feasibility, fidelity, strengths, and challenges of the methods; (d) appropriateness of the coding, scoring, and interpretation of the data; (e) who would serve as the evaluator(s); and (f) the context of the assessment. Moreover, they underscored that each method has strengths and challenges and that no single method can adequately assess the breadth and depth of all competencies. In addition, methods need to be cost-effective and congruent for different stages of professional development as well as for the purpose of assessment (e.g., formative vs. summative evaluation).

MSF or 360-degree evaluation involves the creation of instruments to assess and compare observable behaviors (performance) of professionals. Typically, this involves a self-assessment and an assessment from others (e.g., colleagues, coworkers, clients). An important issue with the use of a questionnaire-based MSF is whether or not it can provide adequate reliability and evidence of validity, demonstrate utility or feasibility, and be applicable to professional psychology. In the following section, we review the use of MSF in the health field addressing many of the issues identified by Kaslow et al. (2009) for its use as an approach for competence-based assessment of psychologists in practice.

The aim of the present study was to conduct a systematic literature review of MSF instruments and to summarize the evidence of feasibility, reliability, generalizability, validity, and other psychometric characteristics of the instruments. Based on the results, we wished to make recommendations about implementing an MSF procedure for assessing psychologists in professional practice.
Method

Data Sources

A systematic literature review was conducted searching for English-language studies published from 1975 to 2012 using the following databases: MEDLINE, EMBASE, CINAHL, PubMed, and PsycINFO. The following terms were used in the search: multisource feedback, 360-degree evaluation, and assessment of medical professionalism.

Studies were included if they (a) described the instrument design; (b) identified factors measured by the instrument; (c) were used with medical/health professionals; (d) included information about feasibility, reliability, generalizability, and validity of the MSF; and (e) were published in English. We excluded studies if (a) they were used with nonhealth/medical professionals; (b) the instruments were not adequately described; (c) they provided only general description and information about MSF; (d) they were published in languages other than English; (e) they provided inadequate information about sample size, study design, data analyses, and psychometric results; and (f) they provided only changes in raters’ performance after feedback.

Data Extraction

Each article in this study was evaluated for inclusion by the third and fourth authors independently based on the title and abstract. Any disagreements were solved by retrieving the full article and having it reviewed by the second author. Based on discussions among the three coders, we achieved 100% agreement for all the included studies.

The initial search yielded 1,061 articles. Of these, 743 articles were excluded based on the title, a further 219 articles were excluded based on the abstract, and another 51 were eliminated after reading the full articles. Finally, we agreed on 48 articles to be included. The detailed characteristics of those studies are described in Table 1.

Implementation Process of Surveys and Rater Selection Across Studies

Candidates selected for MSF (e.g., family physicians, psychiatrists, occupational therapists) are typically provided self-report questionnaires, patient questionnaires (typically 25 questionnaires), coworker questionnaires (typically eight to 10 questionnaires), and colleague questionnaires (typically eight to 10 questionnaires) from the researcher (MSF office). Generally, the candidates or designates (e.g., receptionist) administer questionnaires to the patients, which are then collected and returned to the researcher or MSF office. The candidates distribute questionnaires to his or her coworkers and colleagues to complete, and these raters return the questionnaires to the researcher or MSF office in self-addressed, stamped envelopes. The candidates select raters (e.g., colleagues) who are known to the candidate and who have similar practice (e.g., surgeons for surgeons, etc.).

Results

As shown in Table 1, of 1,061 articles, 48 met the inclusion criteria and 1,013 were excluded. Most studies \( n = 42, 87\% \) were published between 2000 and 2012, another four \( 8\% \) were published between 1990 and 1999, and two \( 5\% \) were published before 1989. Sixteen studies \( 33\% \) were conducted in Canada, 14 \( 29\% \) in the United States, 13 \( 27\% \) in the United Kingdom, one \( 2\% \) in the Netherlands, one \( 2\% \) in Australia, one \( 2\% \) in China, one in Denmark \( 2\% \), and one \( 2\% \) in Taiwan. MSF has been used in almost all the medical specialties but primarily with family physicians and pediatricians. The specific medical and health care specialties were family physicians \( n = 8, 17\% \), obstetrics and gynecology \( n = 3, 6\% \), pediatrics \( n = 5, 10\% \), radiology \( n = 1, 2\% \), and anaesthesia \( n = 3, 6\% \), histopathology \( n = 1, 2\% \), various specialties \( n = 11, 23\% \), psychiatry \( n = 4, 8\% \), occupational therapists \( n = 1, 2\% \), medical radiation technologists \( n = 1, 2\% \), pathologists \( n = 1, 2\% \), emergency medicine \( n = 1, 2\% \), surgery \( n = 4, 8\% \), urology \( n = 1, 2\% \) and internal medicine \( n = 3, 6\% \).

In total, the studies \( 66\% \) used multiple questionnaires to assess candidates, whereas 16 studies \( 34\% \) used a single questionnaire. The domains assessed by MSF included professionalism (reported in 39 \( 81\% \) of total studies), clinical competence \( n = 37, 77\% \), communication \( n = 35, 72\% \), case management \( n = 20, 42\% \), interpersonal relations \( n = 28, 58\% \), and overall assessment \( n = 4, 8\% \).

Description of MSF Instruments

Information about different types of MSF instruments is provided in Table 1. Thirty-two \( 66\% \) studies used multiple surveys in MSF in assessing participants. Most of those studies were conducted in Canada and the United States (see Table 1). The remaining 16 \( 34\% \) studies used single survey in MSF (five used the Sheffield Peer Ratings Assessment Tool, which consists of 24 items). Those studies were mainly conducted in the United Kingdom. The other studies used a single survey with different number of items in each specific survey (see Table 1).

Feasibility

Thirty-nine studies \( 81\% \) addressed feasibility. Most of the studies focused on the response rates for the surveys (see Studies 3–7, 9–13, 15, 17–25, 27–29, 31–36, and 38–48 in Table 1). In general, feasibility was classified as good but the results varied across the studies. Davis (2002), DiMatteo and DiNicola (1981), Lockyer, Violato, Fidler, and Alakija (2009), and Violato, Lockyer, and Fidler (2006) reported response rates of 100% across some of the questionnaires (see Studies 25, 17, 10, 32, respectively, in Table 1). Other researchers reported response rates ranging from 56.2% to 95.1%. One of the studies reported time needed per doctor to complete the questionnaire as a measurement of responses (Lockyer, Violato, & Fidler, 2006a). The authors found that the time required to complete each questionnaire is 6 min, which is considered feasible. In another study, Wood et al. (2006) reported that more than 90% of the 360-degree forms were completed by all raters in less than 1 min, which further supports the feasibility of using such methods.

Internal Structure, Reliability, and Generalizability

Several reliability coefficients are reported in Table 1. Internal consistency reliability (Cronbach’s alpha) was reported as \( \geq 90 \) in
<table>
<thead>
<tr>
<th>Study, country</th>
<th>Specialty (sample size)</th>
<th>Instrument</th>
<th>Raters</th>
<th>Domains assessed</th>
<th>Reliability coefficients</th>
<th>Feasibility</th>
<th>Generalizability</th>
<th>Highest validity evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Violato, Lockyer, &amp; Fidler (2008b), Canada</td>
<td>Family physicians (250)</td>
<td>Medical colleagues, coworkers, patients, self</td>
<td>25 patients, 8 medical colleagues raters (peers and staff mixed)</td>
<td>Professional, clinical competence, communication, manager</td>
<td>α = .96–.98</td>
<td>Response rates = 92.75–96.9%</td>
<td>Ep² = .78–.83</td>
<td>Construct validity</td>
</tr>
<tr>
<td>2. Wood et al. (2006); United Kingdom</td>
<td>Obstetricians and gynecologists (180)</td>
<td>MSF: 4 items</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Construct validity</td>
</tr>
<tr>
<td>3. Archer, McGraw, &amp; Davies (2010); United Kingdom</td>
<td>Pediatrics (577)</td>
<td>MSF (SPRAT): 24 items</td>
<td>8–12 mixed raters</td>
<td>Professional, clinical competence</td>
<td>8 raters; ICC = .80</td>
<td>Response rate = 83%</td>
<td>Construct validity</td>
<td></td>
</tr>
<tr>
<td>4. Wood et al. (2004); United States</td>
<td>Radiology (7)</td>
<td>MSF: 10 items</td>
<td>3 mixed raters</td>
<td>Professional, communication</td>
<td>8 raters; ICC = .85–.87</td>
<td>Response rate = 100%</td>
<td>Content validity</td>
<td></td>
</tr>
<tr>
<td>5. Lockyer, Violato, &amp; Fidler (2007); Canada</td>
<td>Family physicians (250)</td>
<td>MSF: 31 items</td>
<td>1 rater (self-rating)</td>
<td>Professional, clinical competence, communication, manager</td>
<td>α = .74–.98</td>
<td>Response rate = 100%</td>
<td>Construct validity</td>
<td></td>
</tr>
<tr>
<td>6. Meng, David, &amp; Rita (2009); United States</td>
<td>Anesthesia (15)</td>
<td>MSF: 31 items</td>
<td>Mean = 22 mixed raters</td>
<td>Professional, communication</td>
<td>ICC = .82–.87</td>
<td>Response rate = 88%</td>
<td>Content validity</td>
<td></td>
</tr>
<tr>
<td>7. Davies et al. (2008); United Kingdom</td>
<td>Histopathology (92)</td>
<td>MSF (PAT-SPRAT): 24 items</td>
<td>10 raters</td>
<td>Professional, clinical competence, manager, interpersonal relationship</td>
<td>With 8 raters, ICC &gt; .70</td>
<td>Response rate = 92%</td>
<td>Construct validity, factor analysis</td>
<td></td>
</tr>
<tr>
<td>8. Brinkman et al. (2007); United States</td>
<td>Pediatric (36)</td>
<td>MSF: parents = 10 items; nurses = 14 items</td>
<td>Parents and nurses</td>
<td>Professional, clinical competence</td>
<td>Parent α = .95; nurse α = .95</td>
<td></td>
<td>Construct validity, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>9. Archer, Nocini, Southgate, Heard, &amp; Davies (2008); United Kingdom</td>
<td>Various specialties (553)</td>
<td>MSF (Mini-PAT): 16 items</td>
<td>6–8 raters</td>
<td>Professional</td>
<td>Mini-PAT α = .98; ICC &gt; .70</td>
<td>Response rate = 67%</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>10. Davis (2002); United States</td>
<td>Obstetricians and gynecologists</td>
<td>MSF: 16 items</td>
<td>15 peers, 14 attending doctors, 6 nursing and self</td>
<td>Professional, communication, interpersonal overall</td>
<td>ICC = .34–.84</td>
<td>Response rates = 92.9–100%</td>
<td>Criterion-related validity, Pearson’s r</td>
<td></td>
</tr>
<tr>
<td>11. Violato, Marini, Toews, Lockyer, &amp; Fidler (1997); Canada</td>
<td>Family physicians and specialists (28)</td>
<td>Peer, self, patients, coworker, consultant, referring physician</td>
<td>Patients, physician, coworkers, peer, self</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>α = .89–.95</td>
<td>Response rates = 73.5–87.4%</td>
<td>Ep² = .82–.84</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>12. Lockyer &amp; Violato (2004); Canada</td>
<td>Psychiatrists, pediatricians, and internal medicine (304)</td>
<td>MSF: 36 items</td>
<td>8 peer raters</td>
<td>Professional, clinical competence, communication</td>
<td>Peer α = .98</td>
<td>Peer response rate = 94.8%</td>
<td>Ep² = .70–.82</td>
<td>Content validity</td>
</tr>
<tr>
<td>Study; country</td>
<td>Specialty (sample size)</td>
<td>Instrument</td>
<td>Raters</td>
<td>Domains assessed</td>
<td>Reliability coefficients</td>
<td>Feasibility</td>
<td>Generalizability</td>
<td>Highest validity evidence</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>--------</td>
<td>------------------</td>
<td>-------------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>13. Violato, Worsfold, &amp; Polgar (2009); Canada</td>
<td>Occupational therapists (238)</td>
<td>MSF: coworkers, self, and clients</td>
<td>15 clients and 12 coworkers</td>
<td>Professional, clinical competence, communication, manager</td>
<td>$\alpha = 0.97-0.93$</td>
<td>Response rates $= 80.7-95.2%$</td>
<td></td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>14. Lipner et al. (2002); United States</td>
<td>Various specialties (356)</td>
<td>MSF: patient, peer</td>
<td>25 patients, 10 peer, 2 self-ratings</td>
<td>Professional, communication</td>
<td>Between items ranging from $0.43$ to $0.62$</td>
<td></td>
<td>$E_{p}^{2} = 0.61-0.67$</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>15. Violato &amp; Saberton (2006); Canada</td>
<td>Medical radiation technologists (307)</td>
<td>MSF: patients, coworker, colleagues, self</td>
<td>25 patients, 6 colleagues, 6 coworker, and self-ratings</td>
<td>Professional, clinical competence, manager, interpersonal</td>
<td>$\alpha = 0.96-0.97$</td>
<td>Response rates $= 91-100%$</td>
<td></td>
<td>$E_{p}^{2} = 0.78-0.81$</td>
</tr>
<tr>
<td>16. Violato &amp; Lockyer (2006); Canada</td>
<td>Psychiatrists, pediatricians, and internal medicine (304)</td>
<td>MSF: 38 items</td>
<td>Patients, coworkers, psychiatrist colleagues, and self</td>
<td>Professional, clinical competence, communication</td>
<td></td>
<td></td>
<td>$E_{p}^{2} = 0.78-0.82$</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>17. Lockyer et al. (2009); Canada</td>
<td>Pathologists/laboratory medicine (101)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>8 coworkers, 8 peers, and 8 referring physicians</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>$\alpha = 0.95-0.98$</td>
<td>Response rates $= 91-100%$</td>
<td></td>
<td>$E_{p}^{2} = 0.78-0.81$</td>
</tr>
<tr>
<td>18. Violato et al. (2008a); Canada</td>
<td>Psychiatrists (101)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>30 patients, 8 medical colleagues, 8 coworkers, and self-assessment</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>$\alpha = 0.96-0.98$</td>
<td>Response rates $= 92.1-100%$</td>
<td></td>
<td>$E_{p}^{2} = 0.78-0.82$</td>
</tr>
<tr>
<td>19. Lockyer et al. (2006a); Canada</td>
<td>Emergency medicine (187)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>25 patients, 8 medical colleagues, 8 coworkers, and self-assessment</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>$\alpha = 0.68-0.97$</td>
<td>Response rates $= 93-96%$</td>
<td></td>
<td>$E_{p}^{2} = 0.68-0.85$</td>
</tr>
<tr>
<td>20. Lockyer et al. (2006b); Canada</td>
<td>Anesthesiologists (197)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>30 patients, 8 medical colleagues, 8 coworkers, and self-assessment</td>
<td>Professional, communication, interpersonal relationship</td>
<td>$\alpha = 0.93-0.97$</td>
<td>Response rates $= 56.2-95.1$</td>
<td></td>
<td>$E_{p}^{2} = 0.56-0.69$</td>
</tr>
<tr>
<td>21. Violato et al. (2003); Canada</td>
<td>Surgery (252)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>25 patients, 8 medical colleagues, 8 coworkers, and self-assessment</td>
<td>Professional, clinical competence, manager, interpersonal</td>
<td>$\alpha = 0.93-0.97$</td>
<td>Response rates $= 83.2-96.5%$, medical</td>
<td>$E_{p}^{2} &gt; 0.70$ for 8 assessors and 25 patients</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>22. Lockyer, Violato, &amp; Fidler (2003); Canada</td>
<td>Surgery (153)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>Self-evaluation</td>
<td>Professional, communication</td>
<td>$\alpha = 0.78-0.93$</td>
<td>Response rate $= 76.5%$</td>
<td></td>
<td>Construct validity, factor analysis (table continues)</td>
</tr>
<tr>
<td>Study; country</td>
<td>Specialty (sample size)</td>
<td>Instrument</td>
<td>Raters</td>
<td>Domains assessed</td>
<td>Reliability coefficients</td>
<td>Feasibility</td>
<td>Generalizability</td>
<td>Highest validity evidence</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>23. Allerup et al. (2007); Denmark</td>
<td>Internal medicine (42)</td>
<td>MSF: 15 items</td>
<td>1 secretary, 4 nurses, 5 senior doctors</td>
<td>Professional, clinical competence, communication, manager</td>
<td>α = .46–.89</td>
<td>Response rate 95.4%</td>
<td></td>
<td>Criterion-related, Pearson’s r, between-groups differences</td>
</tr>
<tr>
<td>24. Hall et al. (1999); Canada</td>
<td>Family physicians (308)</td>
<td>Self, peer, patients, consultants, referring, coworkers</td>
<td>25 patients, 8 colleagues, 6 coworkers, 6 peers, 6 referrals, self medical colleagues, 8 coworkers, 25 patients, self</td>
<td>Professional, clinical competence, communication, manager</td>
<td>α = .93–.95</td>
<td>Response rates 79.7–88.67%</td>
<td>Ep² = .82</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>25. Violato et al. (2006); Canada</td>
<td>Pediatrics (100)</td>
<td>MSF: patients, coworkers, colleagues, self</td>
<td>8 medical colleagues, 8 coworker, 25 patients, self</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>α = .95–.99</td>
<td>Response rates 93.6–100%</td>
<td>Ep² = .78–.85</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>26. Archer &amp; McAvoy (2011); United Kingdom</td>
<td>Various specialties (68)</td>
<td>SPRAT: 24 items</td>
<td>11 coworkers, 23 patients</td>
<td>Professional, clinical competence, communication</td>
<td>α = .97–.98</td>
<td></td>
<td>Ep² = .75</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>27. Wenrich, Carline, Giles, &amp; Ramsey (1993); United States</td>
<td>Internal medicine (318)</td>
<td>MSF: 13 items</td>
<td>12 nurses</td>
<td>Professional, clinical competence, communication, interpersonal</td>
<td>α = .97–.98</td>
<td>Colleagues M = 12.7; patients M = 19.2</td>
<td></td>
<td>Ep² = .75</td>
</tr>
<tr>
<td>28. LeBlott et al. (2008); United Kingdom</td>
<td>Consultant psychiatrists (347)</td>
<td>MSF: self, colleagues, patients</td>
<td>Self, 15 colleagues, 30 patients</td>
<td>Professional, communication, interpersonal</td>
<td>α = .97–.98</td>
<td></td>
<td>Ep² = .75</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>29. Archer et al. (2005); United Kingdom</td>
<td>Pediatrics (112)</td>
<td>MSF (SPRAT): 24 items</td>
<td>10 mixed (doctors, nurses)</td>
<td>Professional, clinical competence, communication</td>
<td>Four raters ICC &gt; .70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. Campbell, Richards, Greco, Narayanan, &amp; Brearley (2003); United Kingdom</td>
<td>Various specialties (291)</td>
<td>MSF: patients = 18 items, colleagues = 25 items</td>
<td>12 medical colleagues, 30 patients</td>
<td>Professional, clinical competence, communication, manager</td>
<td>α = .89–.92</td>
<td></td>
<td></td>
<td>Ep² = .75–.76</td>
</tr>
<tr>
<td>31. Ramsey et al. (1993); United States</td>
<td>Physicians (313)</td>
<td>MSF: 11 items</td>
<td>11 (physicians and nurses)</td>
<td>Professional, clinical competence, overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. DiMatteo &amp; DiNicola (1981); United States</td>
<td>Various specialties (141)</td>
<td>MSF: physician = 9 items, self = 8 items, patients = 3 items</td>
<td>6 physicians, 10 house staff, 16 patients</td>
<td>Clinical skills, interpersonal relationship</td>
<td>α = .40–.93</td>
<td></td>
<td></td>
<td>Ep² = .70</td>
</tr>
<tr>
<td>33. Noonan, Monagle, &amp; Castanelli (2011); Australia</td>
<td>Consultant anesthetists (60)</td>
<td>MSF: 15 items</td>
<td>1 secretary, 4 nurses, 5 senior doctors</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>α = .46–.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study; country</td>
<td>Specialty (sample size)</td>
<td>Instrument</td>
<td>Raters</td>
<td>Domains assessed</td>
<td>Reliability coefficients</td>
<td>Feasibility</td>
<td>Generalizability</td>
<td>Highest validity evidence</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>--------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>34. Campbell, Narayanan, Burford, &amp; Greco (2010); United Kingdom</td>
<td>General practice (179)</td>
<td>Colleague Feedback Evaluation Tool (CFET) and Doctor’s Interpersonal Skills Questionnaire (DISQ) = CFEP360; colleagues = 18 items, patients = 12 items</td>
<td>1 attending doctor, 2 colleagues, 3 coworkers, 2 office staff, 7 patients, self</td>
<td>Professional, clinical competence, communication, manager, interpersonal</td>
<td>( \alpha = .84-.95 )</td>
<td>Screen for underperformance</td>
<td>Ep(^2) = .80-.81</td>
<td>Construct validity, factor analysis</td>
</tr>
<tr>
<td>35. Qu, Zhao, &amp; Sun (2012); China</td>
<td>Various specialties (258)</td>
<td>Attending doctors = 21 items, colleagues = 21 items, coworkers = 21 items, office staff = 5 items, patients = 25 items, self = 21 items</td>
<td>1 attending doctor, 2 colleagues, 3 coworkers, 2 office staff, 7 patients, self</td>
<td>Professional, communication, clinical competence, manager, interpersonal</td>
<td>( \alpha = .90-.93 )</td>
<td>MSF program by China Medical Board for resident physicians in China</td>
<td></td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>36. Warm, Schauer, Revis, &amp; Boer (2010); United States</td>
<td>Internal medicine (22)</td>
<td>Colleagues, coworkers, attending doctors = 4 items, self = 10 items</td>
<td>1 attending doctor, 2 colleagues, 3 coworkers, 2 office staff, 7 patients, self</td>
<td>Professional, communication, clinical competence, manager, interpersonal</td>
<td>( \alpha = .84-.89 )</td>
<td>MSF as mandatory for internal medicine residents</td>
<td>Ep(^2) = .67-.71</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>37. Lockyer, Blackmore, et al. (2006); Canada</td>
<td>General practice (37)</td>
<td>PAR: self = 21 items, colleagues = 22 items, coworkers = 22 items, patients = 13 items</td>
<td>8 colleagues, 6 coworkers, 9 patients, 25 self</td>
<td>Professional, clinical competence</td>
<td>( \alpha = .88-.97 )</td>
<td>Response rates = 88-100%</td>
<td>Ep(^2) = .67-.71</td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>38. Pollock, Donnelly, Plymale, Stewart, &amp; Vasconez (2007); United States</td>
<td>Plastic surgery (6)</td>
<td>Not applicable (NA): colleagues + coworkers = 60 items</td>
<td>12 colleagues, 28 coworkers</td>
<td>Professional, communication, clinical competence, manager, interpersonal</td>
<td></td>
<td>Evaluates resident performance using accessible colleagues and coworkers</td>
<td></td>
<td>Criterion-related: ( r ) between colleague and coworker</td>
</tr>
<tr>
<td>39. Yang et al. (2011); Taiwan</td>
<td>Various specialties (245)</td>
<td>NA: colleagues = 12 items</td>
<td>5 colleagues</td>
<td>Professional, clinical competence</td>
<td>Colleague ( \alpha = .86 )</td>
<td>Feasible to evaluate Year 1 residents</td>
<td>5 colleague Ep(^2) = .80</td>
<td>Construct validity, factor analysis &amp; between-groups differences</td>
</tr>
<tr>
<td>40. Overeem et al. (2012); Netherlands</td>
<td>Various specialties (146)</td>
<td>NA: colleagues = 33 items, coworkers = 22 items, patients = 18 items</td>
<td>7 colleagues, 7 coworkers, 15 patients</td>
<td>Professional, communication, clinical competence, manager, interpersonal</td>
<td>Coworker ( \alpha = .94-.95 )</td>
<td>MSF system in 23 hospitals</td>
<td></td>
<td>Construct validity, factor analysis, between-groups differences</td>
</tr>
<tr>
<td>Study; country</td>
<td>Specialty (sample size)</td>
<td>Instrument</td>
<td>Raters</td>
<td>Domains assessed</td>
<td>Reliability coefficients</td>
<td>Feasibility</td>
<td>Generalizability</td>
<td>Highest validity evidence</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------</td>
<td>------------</td>
<td>--------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Chandler et al. (2010); United States</td>
<td>Pediatrics (66)</td>
<td>NA: colleagues = 10 items, coworkers = 10 items, patients = 10 items, self = 10 items</td>
<td>3 colleagues, 7 coworkers, 11 patients, 1 self</td>
<td>Professional, communication</td>
<td>Ratings of resident’s professionalism and interpersonal skills</td>
<td>3 nurses, 2 rehab staff, 13 medical students Ep^2 = .70</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Massagli &amp; Carline (2007); United States</td>
<td>Physician and medicine rehabilitation (56)</td>
<td>NA: nurses, medical students, rehab staff</td>
<td>4 nurses, 3 medical students, 10 rehab staff</td>
<td>Professional, clinical competence</td>
<td>Feasibility for physical medicine and rehabilitation residents</td>
<td>3 nurses, 2 rehab staff, 13 medical students Ep^2 = .70</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Risucci et al. (1989); United States</td>
<td>Surgical residents (32)</td>
<td>NA: colleagues + self = 18 items</td>
<td>23 colleagues, 1 supervisor, 27 self</td>
<td>Professional, clinical competence, interpersonal</td>
<td>Differentiates attending surgeons from surgical residents</td>
<td>23 patients Ep^2 = .70</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Sinclair et al. (2009); United Kingdom</td>
<td>Urologist consultants (10)</td>
<td>SHEFFPAT: patients = 13 items</td>
<td>23 patients</td>
<td>Professional, clinical competence, interpersonal</td>
<td>23 raters ICC &gt; .70</td>
<td>Various informants</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Crossley et al. (2008); United Kingdom</td>
<td>Various specialties 137</td>
<td>SPRAT + SHEFFPAT: colleagues = 24 items, patients = 13 items</td>
<td>9 colleagues + coworkers, 15 patients</td>
<td>Clinical competence, interpersonal</td>
<td>ICC = .78</td>
<td>High response rates for instruments</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Joshi, Ling, &amp; Jaeger (2004); United States</td>
<td>Obstetricians and gynecologists (10)</td>
<td>Colleagues = 7 items, coworkers = 53 items, patients = 10 items, student = 12 items</td>
<td>16 Colleagues, 25 coworkers, 10 patients, 12 students</td>
<td>Clinical competence, interpersonal, communication</td>
<td>ICC = .54-.82</td>
<td>Various informants</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Whitehouse, Hassell, Bullock, Wood, &amp; Wall (2007); United Kingdom</td>
<td>Various specialties (171)</td>
<td>TAB: colleague + coworker = 4 items</td>
<td>3 colleagues, 7 coworkers (74% for colleague + coworker)</td>
<td>Interpersonal, communication</td>
<td>Assessment of interpersonal problems in doctors in training</td>
<td>9 colleagues + coworkers Ep^2 = .80</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
</tr>
<tr>
<td>Wall, Singh, Whitehouse, Hassell, &amp; Howes (2012); United Kingdom</td>
<td>Various specialties (834)</td>
<td>TAB: self = 4 items, colleagues + coworkers = 4 items</td>
<td>834 self, 10 colleagues + coworkers</td>
<td>Interpersonal, communication</td>
<td>Used self-assessment, compared raters scores on TAB</td>
<td>Construct validity, factor analysis, between-groups differences</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ICC = intraclass correlation; MSF = multisource feedback; PAT = Peer Assessment Tool; PAR = Physician Achievement Review; SPRAT = Sheffield Peer Review Assessment Tool; SHEFFPAT = Sheffield Patient Assessment Tool; TAB = Team Assessment of Behaviours.
most of the studies. Intraclass correlations (ICCs; correlation between items and overall ratings) were reported in six (12%) studies, with values typically in the high range (>.70) but with the occasional low outlier (ICC = .34; Davis, 2002). Generalizability/reproducibility coefficients (Ep²) were reported for 23 (48%) studies and were adequate with eight or more raters (i.e., Ep² ≥ .70).

Validity

Of the 48 studies that were included in the review, 35 (73%) had results for construct validity, several studies reported evidence of criterion-related validity, and several others reported evidence of content validity.

Content validity. The content validity studies generally focused on determining whether the content that the instrument contained is an adequate sample of the domain it is supposed to represent. Enhancing content validity of instruments can be achieved (sampling of appropriate content and skills) by using a table of specifications based on the list of core competency areas and methods to assess them and having experts systematically review items to ensure that each competency is adequately assessed. Applying this procedure, W. Hall et al. (1999) and Violato, Lockyer, and Fidler (2003), for example, constructed instruments to assess family physicians and surgeons, respectively, in communication skills, interpersonal skills, collegiality, professionalism, and ability to continuously improve. These researchers developed a committee of experts (i.e., physicians, surgeons, psychometric experts) to construct questionnaires of 34 items for medical colleagues, 19 items for coworkers, 33 items for self-assessment, and 39 items for patients. The questionnaires were subsequently sent to physicians and surgeons to provide systematic feedback (a modified Delphi procedure). Questionnaires were edited following the feedback to enhance content validity of the instruments. Overeem et al. (2012) employed a similar procedure to address content validity of MSF instruments adapted for use in the Netherlands.

Criterion-related validity. Several studies reported criterion related-validity evidence by comparing the results of MSF with the results obtained using another assessment method. Criterion-related validity refers to the relationship between scores obtained using the MSF instruments and scores obtained using one or more other instruments or measures. Risucci, Tortolania, and Ward (1989) examined the predictive validity by comparing MSF with the American Board of Surgery in Training Examination (ABSITE). They found a significant correlation between MSF and the ABSITE (r = .58, p < .01). This relationship suggests that as surgeons received higher ratings in MSF, they also received higher rating scores in the ABSITE.

In their MSF study of 356 physicians, Lipner, Blank, Leas, and Fortana (2002) found that the health of the patient was significantly correlated with overall rating (r = .11, p < .001); those in better health tended to rate their doctors higher. Patients who had spent more time under the doctor’s care tended to rate the doctor higher (r = .08, p < .001), and female doctors received higher ratings than did male doctors (r = .17, p < .001). Internal medicine program directors’ ratings were positively correlated with patient ratings of participants’ humanistic qualities (r = .20, p < .05) but not with the program directors’ overall clinical competence ratings.

Crossley et al. (2008) compared the MSF assessment in the form of Non-Technical Skills for Surgeons (NOTSS) with the Procedures Based Assessment (PBA) global summary and Objective Structured Assessment of Technical Skills (OSATS). They found that the NOTSS scores were positively correlated with PBA global summary scores (r = .48, p < .001). Also MSF in the form of NOTSS was positively correlated with the generic part of the OSATS score (r = .51, p < .001). Similarly, Yang et al. (2011), in their work in Taiwan with residents, found significant correlations between medical colleague assessments and scores on objective performance measures (r = .37 and .72, respectively, p < .01). Evidence for criterion-related validity was adduced in several other studies as well (see Table 1).

Construct validity. Evidence for construct validity, which refers to the nature of the psychological construct or characteristic being measured by the instrument, was reported in the majority of studies. Most of the work for construct validity were factor analyses to determine the number of factors in the various instruments, the factors’ variance-accounting properties, their theoretical meaningfulness, and their coherence. Another common analysis investigating construct validity was between-groups differences typically using analyses of variance.

Violato et al. (2003), for example, conducted principal component factor analysis to derive a five-factor solution for the medical colleague questionnaire accounting for 69% of the variance, three factors for the coworker questionnaire accounting for 70.9%, five factors for the patient questionnaire accounting for 73.5%, and four factors for the self-assessment questionnaire accounting for 65.1%. In addition, the mean score was calculated between self-assessment and medical colleague assessment. Surgeons rated themselves lower than did the medical colleagues, with self-assessment M = 4.07 (SD = 0.73) and medical colleague M = 4.50 (SD = 0.64). In a British study of surgical trainees, Crossley et al. (2008) derived four factors with principal component factor analyses in their MSF instruments with six surgical specialties. Overeem et al. (2012) investigated construct validity of MSF instruments adapted for use in the Netherlands with factor analyses. The peer, coworker, and patient instruments, respectively, had six factors, three factors, and one factor with high internal consistencies (Cronbach’s alphas = .95–.96), accounting for 67%, 70%, and 60% of the variance, respectively. They found that peer ratings were positively associated with the patient ratings (r = .214, p < .01) and with coworker ratings (r = .352, p < .01). Coworker ratings were positively associated with patient ratings (r = .220, p < .01).

Although additional work is required to further investigate evidence for the validity of MSF (especially criterion-related, construct, and consequential validity), the research in the present study points to the reliability, feasibility, and tentative validity of MSF and supports its use with health professionals, including its potential use with practicing psychologists.

Domains Assessed by MSF

Professionalism, clinical competence, communication, case management, interpersonal relations, and overall assessment were reported in 81%, 77%, 72%, 42%, 58%, and 8% of the studies, respectively. The term professionalism, however, encompassed several subdomains (psychosocial skills, psychosocial manage-
ment, humanistic qualities, compassion, attitude, professional development, teaching, professional responsibilities, and professional management), and clinical competence included several subdomains (clinical care, good medical practice, patient care, safe practice, clinical performance, knowledge, critical thinking, diagnosis, and management of complex problem). Communication encompassed subdomains as well (communication with staff and interpersonal communication skills), as did case management (reporting, self-management, administrative skills, office personal, access to doctor, practice process, physical office, and physical space). Interpersonal relationships encompassed relationships with patients, colleagues, and family members; collegiality; collaborator; patient education; information provision; and patient interaction. The last factor was overall assessment.

Discussion

A summary of the MSF empirical review indicates some opportunities for professional psychology. First, MSF has been effectively used with many medical and health professionals (e.g., family physicians, nurses, psychiatrists, anesthesiologists, obstetricians, gynecologists, radiologists, occupational therapists) primarily in Canada, the United States, and the United Kingdom to assess their performance on professionalism, interpersonal relations, clinical competence, communication, and case management. Moreover, the surveys that have been developed and used for these evaluations have proven to be appropriate for all of the informants (self, colleague, coworker, and patient) across all the reviewed studies. Hence, it appears that MSF could similarly be appropriately constructed for use with practicing psychologists as well as confidently and effectively used throughout Canada, the United Kingdom, and the United States to assess the performance and competence of psychologists on professionalism, interpersonal skills, clinical competence, communication, and case management. As MSF has been used effectively to assess the competence and performance of medical health professionals in a number of countries, it can be used for the same purposes for psychologists as well.

Second, the accumulated evidence from reported studies from 1980 to 2012 indicates that MSF has adequate to good reliability, validity, and feasibility. The overall internal consistency reliability (Cronbach’s alpha) of MSF instruments is generally greater than .90 for self and other raters such as patients, coworkers, and colleagues. Mean ratings for all MSF instruments are typically from 4 to 5 (when using 5-point Likert scales and an unable to assess category). Generalizability coefficients for the assessors across persons are approximately .80.

There is substantial evidence of content and criterion-related validity and some evidence of construct validity of the MSF instruments applied in the medical professions. Factor analytic studies, for example, have shown that factors tend to be consistent with the intent of the questionnaires, theoretically meaningful, and coherent. Hence, MSF surveys could be constructed for and used by practicing psychologists with similar accumulated evidence of validity, reliability, and feasibility. Although there are issues that need to be further addressed for the use of MSF (e.g., variable reliability across domains), the overall evidence is that the system has very good response rates (i.e., >70%) and is generally considered to be cost-effective. In this regard, the empirical review indicated that MSF can be carried out by a relatively low number of raters (eight to 10 colleagues, coworkers, 25 patients/clients), and that it generally takes these raters about 6 min or less to respond to the items on the questionnaires. Hence, it would seem logical to assume that the use of MSF in the profession of psychology could be reliably, validly, and feasibly done.

Third, respondents report that the MSF system benefits their practice for personal/professional development (e.g., helps them to focus learning activities to legitimate needs) and for multidisciplinary teamwork (Violato & Lockyer, 2006). Based on our review, MSF appears to be useful for both formative (Allerup et al., 2007; Musick, McDowell, Clark, & Salcido, 2003) and summative evaluations (Lockyer & Clyman, 2008) as well as helpful with quality assurance processes and procedures for training programs (Archer, Norcini, & Davies, 2005). MSF could be similarly beneficial for practicing psychologists (i.e., for personal/professional development, formative and summative evaluation, professional quality assurance). It appears that negative feedback from MSF can evoke negative feelings and interfere with its acceptance in some situations. To overcome this possible distress, it may be helpful to provide interventions for professionals that help them focus their feedback on performance tasks and that facilitate their reflection on the feedback. From our review, the research indicates that feedback might be better received if the respondents are familiar with whom they are rating and are able to observe their professional practice (Sargeant et al., 2003). In addition, this feedback needs to be specific, credible, and useful (Sargeant, Mann, Sinclair, van der Vleuten, & Metsemakers, 2008). Researchers contend that MSF can be a positive approach for practice improvement provided that, for example, skilled facilitators are available to encourage reflection, concrete goals are set, and follow-up interviews are carried out (Overeem et al., 2009).

Lastly, it is important to note that a typical barrier in using MSF is the occasional difficulty in recruiting enough coworkers and colleagues to do the ratings in some practice situations. In this regard, a unique barrier for the use of MSF with practicing psychologists as was found with psychiatrists (Violato, Lockyer, & Fidler, 2008a) might be that some patients/clients could be too cognitively and/or emotionally incapacitated (e.g., severely depressed, psychotic, intellectually disabled, etc.) to be able to provide patient/client ratings. Hence, selection and recruitment of patients/clients for some practicing psychologists may be more difficult than for others. Moreover, it is also the case that patient satisfaction ratings within the medical field must be contextualized and interpreted relative to particular patient illnesses and issues; in a similar fashion, it will be important to contextualize the type of patients/clients selected and recruited for the evaluation of practicing psychologists with respect to their clinical diagnoses and associated issues (e.g., attention-deficit/hyperactivity disorder, mood disorder, personality disorder). However, notwithstanding these types of barriers and challenges, MSF is quite flexible and can be constructed and used across many different types of practicing psychologists (e.g., clinical psychologists, counseling psychologists, school psychologists) and with respect to their patients/clients.

In addition, MSF can be adapted for unique practice situations (as it has been for radiology and laboratory medicine in the medical field) in which patients are not directly involved with the professional being assessed. In such cases, including situations like
this in psychology, other instruments (e.g., surveys for referring psychologists) can be developed when patient instruments cannot be employed. In addition, there are particular areas of competence unique to specialties of psychology (e.g., counseling psychology, school psychology) that need to be considered and addressed in the development and use of MSF questionnaires with particular practicing psychologists that have not been addressed in the fields of medicine and business (e.g., the evaluation of a psychologists’ ability to create and maintain a therapeutic alliance with his or her patient/client or a working alliance with allied professionals within the various settings in which he or she works).

**Recommendations for Assessing Psychologists in Practice**

In 2002, a conference titled “Future Directions in Education and Credentialing in Professional Psychology” provided a forum for interorganizational discussion for competence needs in the profession that included members from many associations including the American Psychological Association and the Association of Psychology Postdoctoral and Internship Centers (Kaslow, 2004). The conference resulted in a model of understanding the developmental nature of competence in professional psychology across the professional life span: A cube model of foundational domains (e.g., professionalism, reflective practice/self-assessment, scientific knowledge and methods, relationships, ethical and legal practice, individual and cultural diversity, interdisciplinary systems), functional domains (e.g., assessment, diagnosis, and conceptualization, intervention, consultation, research, evaluation, supervision, training management, administration) of competence, and the stages of professional development of these domains (doctoral education, internship/residency, postdoctoral residency, and continuing competency; Rodolfa et al., 2005). Based on our review of the use of MSF for the assessment of competence of practicing health professionals and the developmental nature of competence in professional psychology, MSF seems well suited for assessing the competence of psychologists in practice.

In 2006, the American Psychological Association Task Force on Assessment of Competency in Professional Psychology provided a report that noted four systems used by health professions to produce competent professionals: entry-level education and training, licensure and practice regulation, continuing professional education, and specialty certification. Four models of assessment were proposed: (a) measures of knowledge (e.g., by way of the Examination for Professional Practice in Psychology), (b) measures of professional decision making, (c) measures of practice performance including professional attributes, and (d) integrated assessments of practice-based skills and tasks (e.g., by way of assessment within education, training, and supervision experiences; American Psychological Association, 2006). Based on our review, it appears that an MSF system can be employed as a measure of performance and competence of practicing psychologists. Licensure and practice regulation of psychologists can potentially involve the assessment of their performance and competence (including professional attitudes) by way of a valid, reliable, and feasible MSF assessment during education, training, and supervision experiences as well. However, this suggestion widens the scope beyond the postlicensure stage of psychologists, and more groundwork will have to be done to review and propose the use of MSF relative to its use with respect to education, training, and supervision. In this regard, this present study and our recommendations have focused on (and have laid the groundwork for) the potential use of MSF with practicing psychologists. Our suggestion would be that MSF profiles be reviewed by members of a psychologists performance committee (PPC), a multimember (e.g., n = 6) regulatory body appointed by the group responsible for administering the program. Should the MSF surveys flag a potential problem, the PPC can work with the psychologist from a quality-improvement perspective. Peer office reviews or other competency assessment tools may be used to assist these psychologists. Confidentiality of information gathered by the questionnaires should be guaranteed under the jurisdiction’s privacy laws. MSF information should be for educational (formative) purposes and should not be used in legal or disciplinary proceedings.

Currently, the primary method used by practicing psychologists to evaluate their performance and competence as well as assess their learning needs is by way of self-assessment (Belar et al., 2001; Pope, Sonne, & Greene, 2006). Self-assessment involves self-reflection and evaluation of one’s professional strengths and areas of improvement in foundational and functional domains as well as an evaluation of one’s limitations and decisions about how to address their developmental needs (Caverzagie, Shea, & Kogan, 2008; Kaslow, Rubin, Forrest, et al., 2007). A major problem with self-assessment as an approach for evaluating competence is that very few self-assessment measures have established adequate psychometric properties, and they tend not to correlate well with ratings by peers and/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, Hall, & Be-resin, 2006).

Systematic reviews in medicine (e.g., Davis et al., 2006) as well as empirical investigations in social psychology (Krueger & Muller, 2002) have suggested that physicians can be limited in their ability accurately self-assess when their assessments are compared with objectively observed and measured competencies. In a meta-analysis of quantitative self-assessment studies in higher education, Falchikov and Boud (1989) reported the results of 44 studies in a variety of subject areas (e.g., medicine, law, engineering, psychology) that showed, on average, that self-assessors were poor to moderate judges of their performance (correlations between self-assessed and external measures of performance ranged from −.05 to .82, with a mean correlation of .39). Hence, it is possible that practicing psychologists may over- or underestimate their performance, abilities, and skills in their self-assessments. Moreover, as suggested by Wise (2010), most psychology registration and licensing boards are unable to ensure that psychologists are maintaining their professional competence, adequately enhancing their professional skills, or remediating their skill deficits by way of psychologist self-assessment.

Although educational training, professional development, ethical guidelines, professional standards, and self-reflection/assessment are necessary aspects of quality assurance for the practice of psychology, they are not sufficient. The primary, or sole, use of self-assessment by practicing psychologists to ensure the quality of education and psychology registration and licensing boards that are providing quality care is insufficient. Accordingly, psychology registration and licensing boards should mandate that the perfor-
mance and competencies of practicing psychologists be reviewed and evaluated by way of an MSF system on some temporal cycle (e.g., every 5 years).

Rodolfa, Schaffer, and Webb (2010) remind us that our field has “operated on the assumption that once a person is evaluated to be competent to practice, that person remains competent throughout his or her entire career” (p. 296). J. Hall and Boucher (2003), however, have cautioned that psychologists with the highest incidence of disciplinary action have been in practice for 11 to 25 years. Professional psychology lags behind other health professionals for monitoring of continuing competence after licensure in the public’s perception (Nutt, 2010). There is a public assumption that health professionals undergo periodic evaluation and assessment of their professional skills (AARP, 2007). Once a psychologist is registered and licensed in a state (in the United States) or province (in Canada), however, no further competence or performance assessment is typically conducted; instead, psychologists may be directed to self-reflect and self-assess their professional development and attend continuing education events to maintain their competence. Moreover, there is no independent evidence that provides support for these efforts in maintaining or developing one’s competence (Rodolfa et al., 2010). Many medical health professionals have taken steps toward assuring the public of their continued competence by developing a framework for maintenance of licensure that includes an ongoing process of self-assessment, self-evaluation, and professional development in deficit areas identified through these assessments as well as their demonstrated competence for patient care, professionalism, and communication skills by way of peer assessment, patient reviews, satisfaction surveys, and MSF (Special Committee on Maintenance of Licensure, 2008). Although medical health professionals appear to be far ahead of psychologists in attending to quality assurance, there have been attempts to move the profession of psychology toward using approaches that more comprehensively and objectively monitor and evaluate the practice of psychology beyond educational and training environments (e.g., Kaslow et al., 2009). To date, psychology registration and licensure boards in Canada and the United States have been not moved beyond self-assessment as a means of competence evaluation of practicing psychologists. Therefore, provincial and state licensing boards for psychology in Canada and the United States, respectively, must undertake greater responsibility beyond their efforts to date (e.g., recommending/mandating periodic self-reflection and self-assessment among licensed psychologists) for ensuring the public that psychologists are practicing in a competent manner because they are the only entity with legal authority over psychologists’ practice (Swankin, LeBuhn, & Morrison, 2006).

Conclusion

Based on empirical evidence from health care and our own professional experience, we propose that an MSF assessment system (that is reliable, valid, and feasible) should be developed to monitor the core competencies of psychologists (for an example of what MSF might look like for school psychologists, see Andrews & Violato, 2010). Moreover, our contention is that such a system cannot only identify strengths and weaknesses of the core competencies of psychologists (i.e., summative assessment), but also provide useful information and feedback for professional development and enrichment (i.e., formative assessment). In addition, an MSF competence-based assessment system for psychologists can provide feedback to psychologists about their performance and improve their practice. An MSF assessment system that incorporates self-assessment along with peer, coworker, and client/patient assessment could provide information that is not only personally/professionally useful for practicing psychologists but useful for national, provincial, and statewide psychology associations as well as provincial and statewide psychology regulatory boards for their oversight and governance of professional psychology.

One step toward incorporating MSF in psychology more broadly is to have one provincial or statewide psychology association and/or regulatory board or association of state and provincial psychology boards support the formation of an MSF advisory group, develop a table of specifications for MSF instruments, and create items from which a pilot study involving practicing psychologists within that jurisdiction could be conducted to provide evidence of the reliability, validity, and feasibility of MSF with practicing psychologists.

References


Received October 26, 2012
Revision received March 12, 2013
Accepted March 14, 2013

---

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://notify.apa.org/ and you will be notified by e-mail when issues of interest to you become available!