

2004

Assessing Mediator Performance: The Usefulness of Participant Questionnaires

Roselle L. Wissler

Robert W. Rack Jr.

Follow this and additional works at: <https://scholarship.law.missouri.edu/jdr>



Part of the [Dispute Resolution and Arbitration Commons](#)

Recommended Citation

Roselle L. Wissler and Robert W. Rack Jr., *Assessing Mediator Performance: The Usefulness of Participant Questionnaires*, 2004 J. Disp. Resol. (2004)

Available at: <https://scholarship.law.missouri.edu/jdr/vol2004/iss1/14>

This Article is brought to you for free and open access by the Law Journals at University of Missouri School of Law Scholarship Repository. It has been accepted for inclusion in Journal of Dispute Resolution by an authorized editor of University of Missouri School of Law Scholarship Repository. For more information, please contact bassettcw@missouri.edu.

Assessing Mediator Performance: The Usefulness of Participant Questionnaires

*Roselle L. Wissler & Robert W. Rack, Jr.**

I. INTRODUCTION.....	229
II. ENSURING QUALITY COURT-CONNECTED MEDIATION	230
III. METHODS FOR ASSESSING MEDIATOR COMPETENCE	232
A. <i>Education, Training, and Experience</i>	232
B. <i>Written Exams</i>	233
C. <i>Settlement Rates</i>	233
D. <i>Performance-Based Assessments</i>	234
E. <i>User Complaints and Assessments</i>	238
IV. THE PRESENT INQUIRY	240
A. <i>The Mediation Process</i>	240
B. <i>The Mediators</i>	241
C. <i>The Survey Procedure and Respondents</i>	242
V. AN EMPIRICAL EXAMINATION OF ATTORNEYS' ASSESSMENTS OF MEDIATOR SKILLFULNESS.....	243
A. <i>Attorneys' Ratings of the Mediators' Skillfulness</i>	243
B. <i>Can the Ratings Differentiate among the Skill Dimensions?</i>	245
C. <i>Can the Ratings Differentiate among the Mediators?</i>	247
D. <i>Do the Ratings Simply Reflect Settlement or Other Mediation Outcomes?</i>	248
E. <i>Are the Ratings Measuring "Skillfulness"?</i>	250
VI. CONCLUSION: THE POTENTIAL USEFULNESS OF PARTICIPANT ASSESSMENTS	253

I. INTRODUCTION

As part of their obligation to provide quality services, courts that offer mediation need to periodically assess the performance of mediators to whom they refer cases.¹ One of several methods that have been proposed for monitoring mediator quality is participant assessments of mediator performance.² The present article reports an empirical study that examined attorneys' assessments of the skillfulness of mediators in a federal appellate civil mediation program.³ The attorneys rated

* Roselle L. Wissler is Director of Research, Lodestar Dispute Resolution Program, Arizona State University College of Law (rwissler@asu.edu). Robert W. Rack, Jr. is Chief Circuit Mediator, U.S. Court of Appeals for the Sixth Circuit (Robert_Rack@ca6.uscourts.gov). The authors thank Wayne Brazil, Bob Dauber, Howard Herman, Chris Honeyman, Craig McEwen, and Ellen Waldman for their comments and Beth DiFelice, Linda Niesen, Wilma Ring, and Kerry Skinner for their assistance.

1. See *infra* Part II.
2. See *infra* Part III.
3. See *infra* Part IV.

some of the mediators as being more skillful than others, and these differences generally remained whether or not favorable outcomes were achieved in mediation. In addition, the attorneys rated individual mediators as being more skillful on some dimensions than others.⁴ These findings suggest that participant assessments could provide an effective means for monitoring mediator performance. We conclude by discussing a number of factors that could affect the usefulness of participant assessments.⁵

II. ENSURING QUALITY COURT-CONNECTED MEDIATION

Mediation programs and mediators have the responsibility to ensure they are providing high quality dispute resolution services.⁶ And courts are responsible for the quality of service provided by any mediators to whom they give their imprimatur—those to whom they refer cases as well as those they employ directly.⁷ Quality control efforts are thought not only to protect consumers but also to maintain the legitimacy of and public confidence in both the mediation process and the courts.⁸ Commentators argue that assuring quality mediation services is particularly important when mediation is funded by public funds, involves mandatory referral or does not permit party choice of the program, the process, or the mediator,⁹ or when the parties would not have the knowledge or the resources to make informed choices.¹⁰

In order to meet their responsibility for ensuring the quality of services provided, courts need to monitor the performance of the mediators as well as the

4. See *infra* Part V.

5. See *infra* Part VI.

6. SPIDR COMMISSION ON QUALIFICATIONS, ENSURING COMPETENCE AND QUALITY IN DISPUTE RESOLUTION PRACTICE, REPORT NO. 2 at 3, 5, 21 (April 1995) [hereinafter SPIDR COMMISSION]; CPR-Georgetown Commission on Ethics and Standards in ADR, *Principles for ADR Provider Organizations*, 56 U. MIAMI L. REV. 983, Principle I., cmt. (2002) [hereinafter CPR-Georgetown Commission]. Judith M. Filner, *Will Mediator Credentialing Assure Quality and Competency?*, DISP. RESOL. MAG., Fall 2001, at 3, 4; Susan S. Silbey, *Mediation Mythology*, 9 NEGOTIATION J. 349 (1993).

7. NATIONAL STANDARDS FOR COURT-CONNECTED MEDIATION PROGRAMS, Standards 2.1, 6.1, 6.3, cmts. (Ctr. for Dispute Settlement & The Inst. of Judicial Admin. 1990) [hereinafter NATIONAL STANDARDS]; SPIDR, QUALIFYING DISPUTE RESOLUTION PRACTITIONERS: GUIDELINES FOR COURT-CONNECTED PROGRAMS iii, 5, 10 (1997) [hereinafter SPIDR, COURT-CONNECTED PROGRAMS].

8. See MARGARET S. HERRMAN ET AL., AN INTERIM REPORT OF THE MEDIATOR SKILLS PROJECT: ASSESSING AND SUPPORTING EFFECTIVE MEDIATION I (1998); NATIONAL STANDARDS, *supra* note 7, at Standard 6.1, cmt.; Juliana Birkoff et al., *Points of View: Is Mediation Really a Profession?*, DISP. RESOL. MAG., Fall 2001, at 10; Wayne D. Brazil, *Comparing Structures for the Delivery of ADR Services by Court: Critical Values and Concerns*, 14 OHIO ST. J. ON DISP. RESOL. 715, 727-28, 738 (1999); Christopher Honeyman, *On Evaluating Mediators*, 6 NEGOTIATION J. 23, 36 (1990) [hereinafter Honeyman, *Evaluating*].

9. NATIONAL STANDARDS, *supra* note 7, at Standard 6.3, cmt.; Margaret Shaw, *Selection, Training, and Qualification of Neutrals*, in NATIONAL SYMPOSIUM ON COURT-CONNECTED DISPUTE RESOLUTION RESEARCH 155, 157 (Susan Keilitz ed., 1994); SPIDR COMMISSION, *supra* note 6, at 5, 12, 19-20; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 10; CPR-Georgetown Commission, *supra* note 6, at Principle I.b., cmt.; Robert Dingwall, *Does "Caveat Emptor" Alone Help Potential Users of Mediation?*, 9 NEGOTIATION J. 331, 331-32 (1993).

10. CPR-Georgetown Commission, *supra* note 6, at Principle I.b., cmt.; Dingwall, *supra* note 9, at 331-32.

operation of the mediation program more broadly.¹¹ Much of the discussion about and efforts to assure mediator competence have focused on the process of selecting mediators.¹² That initial determination of qualifications, however, does not ensure continuing quality mediation.¹³ Some commentators maintain that monitoring mediators' on-going performance "may be equal in importance to the initial selection process" and is one of the "most critical components of any system that seeks to ensure mediator competence."¹⁴ Beyond having minimal continuing education or training requirements, few programs engage in on-going, systematic assessment of their mediators.¹⁵

The periodic assessment of mediators can provide information that can help to maintain quality mediation in several ways. By identifying the mediators' strengths and weaknesses, it could give mediators feedback they would like to have and could use for their own skill-enhancement and professional development.¹⁶ The program administrator could use this information to determine which skill areas should be the focus of additional training or which mediators could benefit from additional supervision or co-mediation,¹⁷ as well as to assign cases to mediators according to their strengths.¹⁸ If additional assistance does not ulti-

11. NATIONAL STANDARDS, *supra* note 7, at Standards 2.1, 6.5, 16.1, 16.2, cmt.; Shaw, *supra* note 9, at 157; SPIDR COMMISSION, *supra* note 6, at 3; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 19-20.

12. CHARLES POU, JR., MEDIATOR QUALITY ASSURANCE: FINAL REPORT TO THE MARYLAND MEDIATOR QUALITY ASSURANCE OVERSIGHT COMMITTEE 2-3 (June 2003), available at http://www.policyconsensus.org/pci/policiestools/QA_MD_Report.pdf (last visited Feb. 16, 2004). See also *infra* Part III.

13. NATIONAL STANDARDS, *supra* note 7, at Standard 6.5, cmt.; POU, *supra* note 12, at 22; Dingwall, *supra* note 9, at 333. See also *infra* notes 24, 28 and accompanying text.

14. NATIONAL STANDARDS, *supra* note 7, at Standards 6.5, 6.6, cmt. See also SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 8, 14, 20; SPIDR COMMISSION, *supra* note 6, at 4, 12; CPR-Georgetown Commission, *supra* note 6, at Principle 1.c.; Steven T. Peluso, *Mediating the Licensing and Certification Labyrinth*, DISP. RESOL. MAG., Fall 2001, at 3, 7, 9.

15. POU, *supra* note 12, at 2-3; Honeyman, *Evaluating*, *supra* note 8, at 24. The structure of the program can affect whether its quality assurance efforts focus on mediator selection or on-going assessment, and different structures pose different problems for achieving quality control. See Brazil, *supra* note 8, at 803-07; Brad Honoroff et al., *Putting Mediation Skills to the Test*, 6 NEGOTIATION J. 37, 39 (1990) (explaining that limited supervisory and training resources required the program to place its emphasis on the selection process).

16. See HERRMAN ET AL., *supra* note 8, at 7; SPIDR COMMISSION, *supra* note 6, at 15, 17; Grace E. D'Alo, *Accountability in Special Education Mediation: Many a Slip 'Twixt Vision and Practice?*, 8 HARV. NEGOT. L. REV. 201, 227 (2003) (reporting that most mediators in a special education mediation program thought observation and feedback would be helpful); Dingwall, *supra* note 9, at 333; Honoroff et al., *supra* note 15, at 37, 41-42; Paul J. Spiegelman, *Certifying Mediators: Using Selection Criteria to Include the Qualified-Lessons from the San Diego Experience*, 30 U.S.F. L. REV. 677, 707 (1996); Susan J. Rogers, *Ten Ways to Work More Effectively with Volunteer Mediators*, 7 NEGOTIATION J. 201, 205 (1991) (reporting that volunteer community mediators wanted more feedback about their skills).

17. HERRMAN ET AL., *supra* note 8, at 8-9; NATIONAL STANDARDS, *supra* note 7, at Standard 6.6, cmt.; Dingwall, *supra* note 9, at 333; Honeyman, *Evaluating*, *supra* note 8, at 24; Ellen Waldman, *Credentialing Approaches: The Slow Movement toward Skills-Based Testing Continues*, DISP. RESOL. MAG., Fall 2001, at 13, 14 [hereinafter Waldman, *Credentialing*].

18. HERRMAN ET AL., *supra* note 8, at 35; SPIDR COMMISSION, *supra* note 6, at 15, 17; Honeyman, *Evaluating*, *supra* note 8, at 24.

mately lead to improved performance, the program administrator could remove that mediator from the roster or could stop referring cases to him or her.¹⁹

III. METHODS FOR ASSESSING MEDIATOR COMPETENCE

In this section, we review several commonly discussed methods for assessing on-going mediator performance. The various methods have different advantages and disadvantages and measure different aspects of mediator competence. Accordingly, the general consensus appears to be that programs should not rely on a single method to assess mediator quality, but should use several of these methods in combination.²⁰

A. Education, Training, and Experience

Hours or type of mediation training, amount of mediation experience and, to a lesser extent, academic degrees or professional licenses are typical qualification requirements mediation programs use to select mediators for entry into their program.²¹ Some programs additionally encourage or require their mediators to periodically attend continuing education and training as a means of on-going quality assurance.²² Despite the fact that it is easy and inexpensive for programs to assess whether mediators have met these requirements, there is a general consensus that they are not adequate proxies for mediator competence because they do not guarantee that the mediator has acquired the needed skills or is maintaining them at an adequate level.²³ The empirical evidence of the relationship between these credentials and mediator performance is mixed.²⁴

19. NATIONAL STANDARDS, *supra* note 7, at Standard 6.6, cmt.; Peter R. Maida, *Rosters and Mediator Quality: What Questions Should We Ask?*, DISP. RESOL. MAG., Fall 2001, at 17, 19.

20. See Shaw, *supra* note 9, at 165; SPIDR COMMISSION, *supra* note 6, at 4, 12; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 8, 14.

21. See SARAH R. COLE ET AL., MEDIATION: LAW, POLICY & PRACTICE § 11:2 App. B (2d ed. 1994 & Supp. 2002) (listing statutes and court rules establishing mediator qualifications in each state); OHIO COMMISSION ON DISPUTE RESOLUTION AND CONFLICT MANAGEMENT, NATIONWIDE SURVEY OF MEDIATOR QUALIFICATION STATUTES AND COURT RULES (2001) (listing statutes establishing mediator qualifications in each state), available at <http://www.state.oh.us/cdr/mediatorqualifications/mediatorqualifications.htm> (last visited Feb. 15, 2004); POU, *supra* note 12, at 10-18 (listing efforts undertaken by professional organizations, courts, and agencies to establish mediator qualifications) (these pages also are available, along with an additional list of state mediator membership organization credentialing activities, at http://www.crnfo.org/mediation-program-managers/pg_5.pdf (last visited Mar. 9, 2004)); SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 16-17; W. Lee Dobbins, *The Debate over Mediator Qualifications: Can They Satisfy the Growing Need to Measure Competence without Barring Entry into the Market?*, 7 U. FLA. J.L. & PUB. POL'Y 95, 105-06 (1995); Filner, *supra* note 6, at 5-6; David A. Hoffman, *Certifying ADR Providers*, BOSTON B.J., March/April 1996, at 9, 23; Maida, *supra* note 19, at 19; Waldman, *Credentialing*, *supra* note 17, at 13-14.

22. See SPIDR COMMISSION, *supra* note 6, at 12; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 15; Jay Folberg et al., *Use of ADR in California Courts: Findings & Proposals*, 26 U.S.F. L. REV. 343, 411 (1992); Maida, *supra* note 19, at 19; Donald T. Weckstein, *Mediator Certification: Why and How*, 30 U.S.F. L. REV. 757, 788 (1996).

23. See NATIONAL STANDARDS, *supra* note 7, at Standards 6.2, 6.5, cmts.; Shaw, *supra* note 9, at 163-64; SPIDR COMMISSION, *supra* note 6, at 12, 18; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 12; Filner, *supra* note 6, at 6; Nancy J. Foster & Joan B. Kelly, *Divorce Mediation: Who Should Be Certified?*, 30 U.S.F. L. REV. 665, 674 (1996); Christopher Honeyman, *A Consensus on Mediators' Qualifications*, 9 NEGOTIATION J. 295, 296 (1993) [hereinafter Honeyman, *Consensus*];

B. Written Exams

Written tests have been suggested as a method to assess mediators' decision-making abilities and knowledge about various aspects of mediation.²⁵ A written exam would be relatively easy and inexpensive to administer and could potentially measure some elements of mediator competence, such as knowledge of procedures, ethics, certain substantive matters, laws regulating mediation and how to handle certain problems that can arise in mediation.²⁶ Some commentators argue that written exams cannot measure the critical interactive skills required in mediation and, thus, should not be the sole basis for determining mediator competency.²⁷ A labor relations program that used a four-hour written exam addressing decision-making abilities reported little relationship between candidates' performance on the exam and their subsequent mediation performance.²⁸

C. Settlement Rates

Settlement rates are often proposed as a means of monitoring mediator quality, but the consensus appears to be that this measure should be used with caution and never as the sole measure of performance for several reasons.²⁹ First, settle-

Craig A. McEwen, *Competence and Quality*, 9 NEGOTIATION J. 317 (1993); Waldman, *Credentialing*, *supra* note 17, at 16. For a discussion of concerns that these criteria could limit diversity among mediators, see Stephanie A. Henning, *A Framework for Developing Mediator Certification Programs*, 4 HARV. NEGOT. L. REV. 189, 198-99 (1999); NATIONAL STANDARDS, *supra* note 7, at Standard 6.1, cmt.; SPIDR COURT-CONNECTED PROGRAMS, *supra* note 7, at 12. See generally Dobbins, *supra* note 21.

24. See, e.g., Will Hathaway & Karen Sontag, *Massachusetts Office of Dispute Resolution Superior Court Mediator Qualification Process*, 5 WORLD ARB. & MEDIATION REP. 15, 16 (1994) (noting that mediation training and experience did not seem to be related to performance in a simulated mediation); Christopher Honeyman, *Five Elements of Mediation*, 4 NEGOTIATION J. 149, 159 (1988) (reporting little relationship between prior mediation experience or law school grades and mediator performance) [hereinafter Honeyman, *Five Elements*]; Honoroff et al., *supra* note 15, at 40 (reporting that educational background, amount of mediation training and experience did not consistently predict success in a role-play performance evaluation, which was related to subsequent mediator performance); Edward Kruk, *Practice Issues, Strategies, and Models: The Current State of Family Mediation*, 36 FAM. & CONCILIATION CTS. REV. 195, 199 (1998) (reporting that professional background, mediation experience and training appeared to be related to settlement rates); Roselle L. Wissler, *Court-Connected Mediation in General Civil Cases: What We Know from Empirical Research*, 17 OHIO ST. J. ON DISP. RESOL. 641, 678 (2002) (reporting no relationship between, on the one hand, hours of mediation training or whether training involved role play and, on the other hand, settlement or parties' and attorneys' assessments of the fairness of the mediation process; prior mediation experience, however, was related to greater likelihood of settlement, but not to perceptions of fairness).

25. E.g., SPIDR COMMISSION, *supra* note 6, at 12; Dobbins, *supra* note 21, at 102-03; Margaret S. Herrman et al., *Supporting Accountability in the Field of Mediation*, 18 NEGOTIATION J. 29, 44 (2002) (reporting the development of a written knowledge exam); Peluso, *supra* note 14, at 9.

26. See, e.g., SPIDR COMMISSION, *supra* note 6, at 12; Dobbins, *supra* note 21, at 102-03; Peluso, *supra* note 14, at 9.

27. See SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 14; Teresa V. Carey, *Credentialing for Mediators—To Be or Not to Be?*, 30 U.S.F. L. REV. 635, 642 (1996); Honeyman, *Five Elements*, *supra* note 24, at 158.

28. Honeyman, *Five Elements*, *supra* note 24, at 159.

29. See NATIONAL STANDARDS, *supra* note 7, at Standard 6.5; Shaw, *supra* note 9, at 165; SPIDR COMMISSION, *supra* note 6, at 12; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 15; Dobbins, *supra* note 21, at 109; Folberg et al., *supra* note 22, at 404 (reporting that California judges and

ment is only one possible outcome of mediation and is not necessarily the most important goal that the parties or the program want to achieve.³⁰ Second, non-settlement does not mean that the mediator did not make a substantial contribution to the case, perhaps even helping the parties to resolve many of the issues,³¹ and settlement does not guarantee that all underlying issues have been addressed and permanently resolved.³² Third, the nature of the process by which the settlement was achieved, such as whether it permitted party self-determination, is also considered important in determining mediator quality.³³ Fourth, heavy reliance on settlement as an assessment tool could lead mediators to coerce settlements.³⁴ Finally, making comparisons of settlement rates among mediators within a program would be misleading if some mediators were routinely assigned more difficult cases.³⁵

D. Performance-Based Assessments

Most commentators believe that the best approach for evaluating mediator competence, and one that should be the primary component in every program's efforts to monitor mediator quality, is performance-based assessment.³⁶ Compared to the other methods discussed above, observing the mediator's performance in a simulated or real case is thought to permit a more accurate determination

ADR providers said that settlement rates were one of the two most important criteria for evaluating ADR providers).

30. See SPIDR COMMISSION, *supra* note 6, at 12; Frank E.A. Sander, *The Obsession with Settlement Rates*, 11 NEGOTIATION J. 329, 331 (1995). Other goals might include improving communication and the relationship between the parties; engaging the parties in collaborative problem solving so they can reach a better resolution; providing objective feedback about the issues so the parties can make more informed decisions; increasing party participation in the resolution of their dispute; providing the opportunity for personal transformation; or giving parties the opportunity to explore settlement and to assess whether or not it is possible. See also SPIDR COMMISSION, *supra* note 6, at 5, 14; Robert A. Baruch Bush, *Defining Qualities in Dispute Resolution: Taxonomies and Anti-Taxonomies of Quality Arguments*, 66 DENV. U. L. REV. 335, 349-50 (1989); Tom R. Tyler, *The Quality of Dispute Resolution Procedures and Outcomes: Measurement Problems and Possibilities*, 66 DENV. U. L. REV. 419, 427-30 (1989).

31. See Folberg et al., *supra* note 22, at 404; Sander, *supra* note 30, at 329-30. The program would need to decide how it wants to assess partial settlements and settlements that occur after mediation but before trial. See Shaw, *supra* note 9, at 171 n.63.

32. Shaw, *supra* note 9, at 165; Dobbins, *supra* note 21, at 109; Honeyman, *Evaluating*, *supra* note 8, at 25; Sander, *supra* note 30, at 330-31.

33. See Baruch Bush, *supra* note 30, at 349; Tom R. Tyler & E. Allan Lind, *Procedural Justice*, in HANDBOOK OF JUSTICE RESEARCH IN LAW 65, 74-76 (Joseph Sanders & V. Lee Hamilton eds., 2000).

34. See SPIDR COMMISSION, *supra* note 6, at 12; Wayne D. Brazil, *Court ADR 25 Years after Pound: Have We Found a Better Way?*, 18 OHIO ST. J. ON DISP. RESOL. 93, 121-22 (2002); Honeyman, *Evaluating*, *supra* note 8, at 25; Sander, *supra* note 30, at 330.

35. Shaw, *supra* note 9, at 165; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 15; Honeyman, *Evaluating*, *supra* note 8, at 25.

36. See HERRMAN ET AL., *supra* note 8, at 8; NATIONAL STANDARDS, *supra* note 7, at Standard 6.5, cmt.; POU, *supra* note 12, at 25; Shaw, *supra* note 9, at 165; SPIDR COMMISSION, *supra* note 6, at 3, 12; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 15, 17; THE TEST DESIGN PROJECT, PERFORMANCE-BASED ASSESSMENT: A METHODOLOGY FOR USE IN SELECTING, TRAINING, AND EVALUATING MEDIATORS 7 (NIDR, 1995), available at www.convenor.com/madison/method.pdf (last visited Feb. 15, 2004); Dingwall, *supra* note 9, at 332; Dobbins, *supra* note 21, at 99-101; Folberg et al., *supra* note 22, at 411; Waldman, *Credentialing*, *supra* note 17, at 16; Weckstein, *supra* note 22, at 786.

of quality because it involves the direct observation and evaluation of the mediator's skills in conducting a session.³⁷ To date, performance-based assessments reported in the literature have primarily been used as part of a selection, training, or credentialing process³⁸ rather than to assess on-going mediator performance.³⁹

Most of the programs that use performance-based assessment use an assessment instrument that lists specific dimensions on which the mediators should be evaluated, as well as instructions, criteria, and a scale for numerically rating their performance on each dimension.⁴⁰ The likely advantage of using an instrument that involves making separate assessments of the mediators' level of skill on each of several dimensions is that it should encourage more focused observations and more specific, objective evaluations.⁴¹ Some commentators caution, however, that rating scales cannot entirely remove the subjectivity of the evaluator's judgments⁴² and that some important but less tangible aspects of mediator performance

37. See SPIDR COMMISSION, *supra* note 6, at 12; Dingwall, *supra* note 9, at 332; Folberg et al., *supra* note 22, at 378 (reporting that many ADR providers surveyed believed that competency can best be assessed by evaluating actual performance); McEwen, *supra* note 23, at 317; Waldman, *Credentialing*, *supra* note 17, at 16.

38. See JUDICIAL COUNCIL OF VIRGINIA, GUIDELINES FOR THE TRAINING AND CERTIFICATION OF COURT-REFERRED MEDIATORS Part D (March 1999), available at <http://www.courts.state.va.us/tom.htm> (last visited Feb. 15, 2004); TEST DESIGN PROJECT, *supra* note 36, at 3, 7; Barbara Filner & Michael Jenkins, *Performance-Based Evaluation of Mediators: The San Diego Mediation Center's Experience*, 30 U.S.F. L. REV. 647, 653, 658, 663 (1996); George H. Friedman & Allan D. Silberman, *A Useful Tool for Evaluating Potential Mediators*, 9 NEGOTIATION J. 313, 314 (1993); Hathaway & Sontag, *supra* note 24, at 16; Honeyman, *Five Elements*, *supra* note 24, at 158-59; Honoroff et al., *supra* note 15, at 38-39; Spiegelman, *supra* note 16, at 704-05. See also POU, *supra* note 12, at 10-18; Louise Phipps Senft, *Portrait of a Good Mediator*, 36 MD. B.J., July/August 2003, at 8, 13; Waldman, *Credentialing*, *supra* note 17, at 14-15.

39. See D'Alo, *supra* note 16, at 225.

40. See, e.g., OFFICE OF THE EXECUTIVE SECRETARY OF THE SUPREME COURT OF VIRGINIA, MENTEE EVALUATION FORM (FORM ADR-1001), available at <http://www.courts.state.va.us/forms/home.html> (last visited Feb. 15, 2004); TEST DESIGN PROJECT, *supra* note 36, at 20-30; D'Alo, *supra* note 16, at 226-27, App. B; Filner & Jenkins, *supra* note 38, at 659-63; Hathaway & Sontag, *supra* note 24, at 16; Honeyman, *Evaluating*, *supra* note 8, at 27-30; Honeyman, *Five Elements*, *supra* note 24, at 158-59; Honoroff et al., *supra* note 15, at 39, 41-45; Spiegelman, *supra* note 16, at 704-06; Waldman, *Credentialing*, *supra* note 17, at 14-15. *But see*, e.g., Friedman & Silberman, *supra* note 38, at 314. Most of these programs have experienced staff or outside mediators conduct the assessments. Honeyman suggests that while mediator self-assessment or peer assessment could be helpful for continuing mediator development and training, evaluation by a program administrator or an outside independent consultant is better able to meet the requirements of program quality control. See Honeyman, *Evaluating*, *supra* note 8, at 32-35. For a discussion of options and issues regarding who should conduct the assessments, see NATIONAL STANDARDS, *supra* note 7, at Standard 6.5, cmt.; SPIDR COMMISSION, *supra* note 6, at 3-4, 12; SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 15; POU, *supra* note 12, at 30.

41. See Honeyman, *Evaluating*, *supra* note 8, at 26, 27.

42. See Dingwall, *supra* note 9, at 332; Spiegelman, *supra* note 16, at 707; Weckstein, *supra* note 22, at 785. For a discussion about the possible implications of mediation orientation for mediator performance assessment, see HERRMAN ET AL., *supra* note 8, at 25, 28-32; Filner & Jenkins, *supra* note 38, at 658; McEwen, *supra* note 23, at 319; Spiegelman, *supra* note 16, at 708; Ellen A. Waldman, *The Challenge of Certification: How to Ensure Mediator Competence While Preserving Diversity*, 30 U.S.F. L. REV. 723, 752-53 (1996) [hereinafter Waldman, *Challenge*]. *But see*, e.g., D'Alo, *supra* note 16, at 225-26 (noting that an effort was made to develop rating scales that would accommodate different mediation styles); Honoroff et al., *supra* note 15, at 40, 42 (reporting that evaluators from different fields and backgrounds were able to disregard their own and the mediators' stylistic differences in their ratings).

can be difficult to quantify.⁴³ While there has been much discussion and debate about which skills indicate general mediator competence,⁴⁴ there appears to be agreement that assessment instruments should be program-specific, because different mediation programs in different contexts have different definitions of success and require different sets of skills.⁴⁵

Most of the programs have used simulated cases to assess mediator performance,⁴⁶ but a few programs have assessed mediator performance in actual cases.⁴⁷ The main advantage of simulations is that by using the same case scenario and instructing the actors to give the same responses in each session, all of the mediators would be exposed to essentially the same case, permitting greater comparability across mediators.⁴⁸ One of the disadvantages of simulations is that the session typically is time-limited,⁴⁹ which could make it difficult to evaluate some areas of mediator competence, such as the ability to generate options.⁵⁰ Another limitation of simulations is that actors are likely to present a more restricted range of emotions than actual parties would, making it difficult to accurately evaluate the mediators' skill in dealing with emotional outbursts and overcoming impasses.⁵¹

43. See Daniel Bowling & David Hoffman, *Bringing Peace into the Room: The Personal Qualities of the Mediator and Their Impact on the Negotiation*, 16 NEGOTIATION J. 5, 11-14 (2000); Deborah M. Kolb & Jonathan E. Kolb, *All the Mediators in the Garden*, 9 NEGOTIATION J. 335 (1993); Carrie Menkel-Meadow, *Measuring Both the Art and Science of Mediation*, 9 NEGOTIATION J. 321, 324 (1993).

44. See SPIDR, COURT-CONNECTED PROGRAMS, *supra* note 7, at 17; TEST DESIGN PROJECT, *supra* note 36, at 14-19; Robert A. Baruch Bush, *Mixed Messages in the Interim Guidelines*, 9 NEGOTIATION J. 341, 342 (1993); Honeyman, *Consensus*, *supra* note 23, at 302-06; Honeyman, *Five Elements*, *supra* note 24, at 149-50; Kolb & Kolb, *supra* note 43, at 335; McEwen, *supra* note 23, at 317-18; Menkel-Meadow, *supra* note 43, at 322-23; Richard A. Salem, *The "Interim Guidelines" Need a Broader Perspective*, 9 NEGOTIATION J. 309, 310-11 (1993); Waldman, *Credentialing*, *supra* note 17, at 16, 21. For a listing of skills thought to be necessary for general mediation competence see, for example, HERRMAN ET AL., *supra* note 8, at App.; SPIDR COMMISSION, *supra* note 6, at 11; Margaret S. Herrman et al., *Defining Mediator Knowledge and Skills*, 17 NEGOTIATION J. 139, 141-51 (2001); Honeyman, *Five Elements*, *supra* note 24, at 152-55.

45. SPIDR COMMISSION, *supra* note 6, at 1, 2, 5; TEST DESIGN PROJECT, *supra* note 36, at 3, 20-30 (showing three variants of sample rating scales); Brazil, *supra* note 8, at 726-27; Honeyman, *Evaluating*, *supra* note 8, at 27; Honoroff et al., *supra* note 15, at 42, 46 (noting the added benefits of a program having to explicitly discuss which skills it wants to emphasize); Salem, *supra* note 44, at 309. For a listing of the dimensions and criteria that programs have included in their performance-based assessments, see *supra* note 40.

46. See TEST DESIGN PROJECT, *supra* note 36, at 15, App. B; Filner & Jenkins, *supra* note 38, at 658; Friedman & Silberman, *supra* note 38, at 314; Hathaway & Sontag, *supra* note 24, at 16; Honeyman, *Five Elements*, *supra* note 24, at 158-59; Honoroff et al., *supra* note 15, at 38-39; Senft, *supra* note 38, at 13; Spiegelman, *supra* note 16, at 704.

47. See, e.g., JUDICIAL COUNCIL OF VIRGINIA, *supra* note 38, at D; D'Alo, *supra* note 16, at 227-28.

48. See TEST DESIGN PROJECT, *supra* note 36, at 35; Filner & Jenkins, *supra* note 38, at 658.

49. Simulations generally have been limited to between forty and seventy-five minutes. See Hathaway & Sontag, *supra* note 24, at 16; Honeyman, *Five Elements*, *supra* note 24, at 158; Honoroff et al., *supra* note 15, at 39; Spiegelman, *supra* note 16, at 705.

50. See TEST DESIGN PROJECT, *supra* note 36, at 35; David E. Matz, *Some Advice for Mediator Evaluators*, 9 NEGOTIATION J. 327, 328-29 (1993). To deal with this limitation, some programs have mediator candidates assume that the simulated mediation is the second session rather than the initial one. See also Hathaway & Sontag, *supra* note 24, at 16; TEST DESIGN PROJECT, *supra* note 36, at 35.

51. See Matz, *supra* note 50, at 328-29; Kolb & Kolb, *supra* note 43, at 336. The artificially constructed situation not only can fail to elicit the mediators' "potential repertoires of behaviors," but also can lead them to engage in actions that they would not undertake in a real case in an effort to try to uncover issues they think are hidden in the simulation. *Id.*

The programs in which several experienced mediators evaluated each mediator candidate reported generally consistent ratings for most candidates,⁵² at least when the evaluators had been thoroughly trained.⁵³ The scores on the performance-based assessment appeared to distinguish among the applicants⁵⁴ and to be related to subsequent on-the-job performance.⁵⁵ Because programs have tended to use simulated cases that involved simple fact patterns,⁵⁶ however, the reliability and validity of performance-based assessment instruments in more complicated simulations or in actual cases have yet to be determined.⁵⁷ Another unexplored question is how closely the skill dimensions and rating scales devised by experienced mediators correspond to what matters to participants when assessing the competence or quality of the mediator.⁵⁸

The main disadvantage of performance-based assessment is the difficulty and expense involved in developing and validating the test instrument and the simulations, and the time and cost involved in administering the assessment.⁵⁹ In addition, evaluators sometimes feel unsure how to rate the mediator based on what he or she did or said without knowing the mediator's intention, and find it necessary to inquire about the mediator's strategy after the role play and before finalizing the ratings.⁶⁰

52. See TEST DESIGN PROJECT, *supra* note 36, at 34; Filner & Jenkins, *supra* note 38, at 659; Hathaway & Sontag, *supra* note 24, at 16; Honoroff et al., *supra* note 15, at 40. *But see* TEST DESIGN PROJECT, *supra* note 36, at 38 (suggesting that evaluators discuss a candidate's performance before rating her or him because different evaluators tend to remember different aspects of the mediator's performance).

53. See TEST DESIGN PROJECT, *supra* note 36, at 33, 35 (stating that evaluators had to have at least a half-day training to produce consistent ratings). *But see* Menkel-Meadow, *supra* note 43, at 324 (noting that programs need to determine that consistency in scoring reflects that the evaluation criteria have independent validity and not simply that the evaluators received the same training).

54. See Honeyman, *Five Elements*, *supra* note 24, at 159; Honoroff et al., *supra* note 15, at 39.

55. See, e.g., Friedman & Silberman, *supra* note 38, at 314 (reporting role plays to be "an effective method for assessing the potential of future mediators"); Christopher Honeyman, *The Common Core of Mediation*, 8 MED. Q. 73, 75 (1990) (noting that performance-based assessments were "useful" in selecting and evaluating mediators) [hereinafter Honeyman, *Common Core*]; Honoroff et al., *supra* note 15, at 40 (reporting that candidates with the highest scores on the performance evaluation were "doing very well," while those with somewhat lower scores had encountered "somewhat more problems"). These articles did not specify how the mediators' subsequent on-the-job performance was assessed.

56. See Filner & Jenkins, *supra* note 38, at 658; Friedman & Silberman, *supra* note 38, at 314; Spiegelman, *supra* note 16, at 708.

57. See Menkel-Meadow, *supra* note 43, at 324. The instruments also need to be tested for bias. *Id.* Examination of this issue to date has found no evidence of race or gender bias. See TEST DESIGN PROJECT, *supra* note 36, at 34; CHRISTOPHER HONEYMAN ET AL., IN THE MIND'S EYE? CONSISTENCY AND VARIATION IN EVALUATING MEDIATORS 17-18 (Program on Negotiation at Harvard Law School, Working Paper No. 90-21, 1990).

58. See McEwen, *supra* note 23, at 319.

59. See Dingwall, *supra* note 9, at 332; Honeyman, *Consensus*, *supra* note 23, at 296; Honeyman, *Five Elements*, *supra* note 24, at 159; Spiegelman, *supra* note 16, at 707. To make the use of a performance-based assessment more cost effective, the Test Design Project set out to create standardized evaluation scales that programs could adapt to their specific setting. See TEST DESIGN PROJECT, *supra* note 36, at 9, App. B. Part of the difficulty involves reaching agreement on what skills constitute "quality mediation." See *supra* notes 42, 44 and accompanying text. See also *supra* notes 49-51 and accompanying text (discussing disadvantages specific to the use of simulations in performance-based assessments).

60. See TEST DESIGN PROJECT, *supra* note 36, at App. B; Honeyman, *Evaluating*, *supra* note 8, at 33; Honoroff et al., *supra* note 15, at 38-39, 45; Matz, *supra* note 50, at 328; Spiegelman, *supra* note

E. User Complaints and Assessments

Getting input from users is frequently suggested as one of the ways to evaluate on-going mediator performance for quality.⁶¹ One recommended source of user input includes the review of complaints.⁶² Because complaints tend to be rare,⁶³ and their absence does not necessarily indicate the absence of problems,⁶⁴ complaints are not likely to provide an adequate source of information for on-going quality monitoring.

The other recommended source of user input is participants' assessments and perceptions of the mediators.⁶⁵ Some commentators have noted that participants are uniquely able to evaluate certain aspects of mediator performance.⁶⁶ Interestingly, evaluators in performance-based assessments relied in part on the actors in the simulation when rating the mediators—they observed how the actors reacted to the mediator or asked them after the simulation for their views of the mediator.⁶⁷ A number of programs rely on post-mediation user questionnaires to evaluate the program and the neutrals, although they do not necessarily use that feedback to assess the performance of individual mediators.⁶⁸

16, at 706. *But see* Matz, *supra* note 50, at 328 (noting that post-simulation questioning of the mediator is not always helpful, as the mediators might not remember their strategies and intentions or might make it up).

61. *See* HERRMAN ET AL., *supra* note 8, at 8; NATIONAL STANDARDS, *supra* note 7, at Standard 6.5, cmt.; POU, *supra* note 12, at 4; Shaw, *supra* note 9, at 165; SPIDR COMMISSION, *supra* note 6, at 4, 12; CPR-Georgetown Commission, *supra* note 6, at Principle VI; Folberg et al., *supra* note 22, at 404.

62. *See, e.g.*, POU, *supra* note 12, at 4; Shaw, *supra* note 9, at 165; SPIDR COMMISSION, *supra* note 6, at 4, 12; CPR-Georgetown Commission, *supra* note 6, at Principle VI; Peluso, *supra* note 14, at 9.

63. *See* Jay Folberg, *Certification of Mediators in California: An Introduction*, 30 U.S.F. L. REV. 609 (1996); Robert W. Rack, Jr., *Thoughts of a Chief Circuit Mediator on Federal Court-Annexed Mediation*, 17 OHIO ST. J. ON DISP. RESOL. 609, 625 (2002) [hereinafter Rack, *Thoughts*]; Personal communication with Hon. Wayne D. Brazil, U.S. Magistrate Judge, Northern District of California (Sept. 29, 2003) (notes on file with authors).

64. *See* Henning, *supra* note 23, at 192 (suggesting that a low volume of complaints could reflect inadequacies in the program's grievance procedure or users' uncertainty, due to their unfamiliarity with mediation, about whether there was a problem); Rack, *Thoughts*, *supra* note 63, at 625. Even participants who are familiar with mediation, like many attorneys, might be reluctant to complain about mediators in court-connected programs because they do not want to risk alienating the judges. The presence of complaints, however, probably does signal a problem. *See id.*

65. TIMOTHY HEDEEN, CADRE, USING PARTICIPANT FEEDBACK TO EVALUATE AND IMPROVE QUALITY IN MEDIATION (2002), available at http://www.directionservice.org/cadre/participant_feedback2002.cfm (last visited Feb. 15, 2004); HERRMAN ET AL., *supra* note 8, at 8; NATIONAL STANDARDS, *supra* note 7, at Standard 6.5, cmt.; POU, *supra* note 12, at 4; Shaw, *supra* note 9, at 165; SPIDR COURT-CONNECTED PROGRAMS, *supra* note 7, at 15, 20; SPIDR COMMISSION, *supra* note 6, at 12; Folberg et al., *supra* note 22, at 404 (reporting that California judges and ADR providers said that client satisfaction was one of the two most important criteria for evaluating ADR providers).

66. *See* McEwen, *supra* note 23, at 318-19; Paul G. Garrity, *Certification: What's the Point?*, 40 BOSTON B.J. 8 (1996). *See also* Waldman, *Credentialing*, *supra* note 17, at 16 (noting that performance exams enlist raters as proxies for the actual parties).

67. *See* Filner & Jenkins, *supra* note 38, at 663; Honoroff et al., *supra* note 15, at 42-43; Matz, *supra* note 50, at 328-29.

68. *See* POU, *supra* note 12, at 32, 41; DONNA STIENSTRA ET AL., REPORT TO THE JUDICIAL CONFERENCE COMMITTEE ON COURT ADMINISTRATION AND CASE MANAGEMENT: A STUDY OF THE FIVE DEMONSTRATION PROGRAMS ESTABLISHED UNDER THE CIVIL JUSTICE REFORM ACT OF 1990 255-83 (1997); LEILA TAAFFE & SHINJI MOROKUMA, QUICK SUMMARY OF FINDINGS FROM THE PARTICIPANT SATISFACTION SURVEY OF GEORGIA'S COURT-CONNECTED ADR PROGRAMS 1, avail-

Information obtained from user assessments and aggregated across a large number of cases would be based on a much wider sample of mediator performance than performance-based assessments,⁶⁹ and accordingly could provide a broader and more representative picture of mediator performance across time and cases. For those mediation programs that already use participant questionnaires to monitor program performance, there would be little cost involved in adding specific questions about the mediators or in assessing the performance of individual mediators. Information obtained in one program that had ratings of the mediators by both participants and staff seemed to suggest discrepancies between the two sets of ratings.⁷⁰ However, due to the lack of comparability between the information obtained from these two groups, we cannot draw any conclusions about the relative utility of these two methods.

Questions have been raised about how useful participants' assessments will be for monitoring mediator performance. For example, will participants know enough about what they should expect from a competent mediator to give meaningful assessments and will they take the time to give thoughtful responses?⁷¹ Will participants' assessments be sufficiently sensitive and fine-grained to reveal differences among the mediators on different skills?⁷² Will participants' assessments be biased by their perception that the mediator helped or hurt their case⁷³ or by whether or not the case was settled?⁷⁴ Can participants distinguish between the mediator's performance and other aspects of the process unrelated to competence when making their assessments?⁷⁵

able at <http://www.state.ga.us/gadr/pdfs/sjishort.pdf> (last visited Feb. 15, 2004); *Mediation Program Symposium on USPS REDRESS[®] Program*, 2 CARDOZO ONLINE J. CONFLICT RESOL. (2001), available at http://www.cardozo.yu.edu/cojcr/final_site/symposia/vol_2_symposia/postal_trans (last visited Feb. 15, 2004) (noting the data are not analyzed by individual mediators); Wissler, *supra* note 24, at 645-46, 661-63. *But see, e.g.*, OFFICE OF THE EXECUTIVE SECRETARY OF THE SUPREME COURT OF VIRGINIA, CLIENT EVALUATION OF MEDIATION AND MEDIATORS, PROCEDURES FOR COMPLETION OF CLIENT EVALUATION OF MEDIATION AND MEDIATORS (describing interventions in response to negative evaluations), available at <http://www.courts.state.va.us/forms/home.html> (last visited Feb. 15, 2004); Hathaway & Sontag, *supra* note 24, at 17 (reporting the use of written evaluations by attorneys to monitor on-going mediator performance); Rack, *Thoughts, supra* note 63, at 622; Personal communication with Hon. Wayne D. Brazil, *supra* note 63.

69. See Matz, *supra* note 50, at 328-29 (noting a limitation of performance-based assessment is that the evaluator is judging the mediator in a single case).

70. This program had approximately 2,000 participant questionnaires completed over several years and observations of eight mediation sessions. Although the participant assessments were highly favorable, the observations revealed mediator actions that were not consistent with the program's policies or idea of quality mediation. The eight mediation sessions observed represented a tiny fraction of the program's caseload, however, and there often was not a good fit between the rating categories used in the observations and the questions asked of participants (e.g., the author compared the observation ratings of "moved parties toward new relationship" with participant responses to questions about their degree of trust in the mediator and confidence in the mediator to help resolve the dispute). See D'Alo, *supra* note 16, at 222, 229-32, 236-37, App. A.

71. See Honeyman, *Evaluating, supra* note 8, at 25, 33.

72. Participants' evaluations of the mediator and the mediation process, as reported in studies of mediation program effectiveness, tend to be relatively uniform and highly favorable. See STIENSTRA ET AL., *supra* note 68, at 217, 268, 279; Wissler, *supra* note 24, at 661-63.

73. See Honeyman, *Evaluating, supra* note 8, at 25, 33.

74. See, e.g., Wissler, *supra* note 24, at 661 n.77, 663 n.80 (reporting that parties and attorneys in cases that settled in mediation generally had more favorable assessments of the mediator and the mediation process than participants in cases that did not settle).

75. See SPIDR COURT-CONNECTED PROGRAMS, *supra* note 7, at 15.

IV. THE PRESENT INQUIRY

Despite frequent suggestions that participant assessments be used to monitor on-going mediator performance, many questions about their potential usefulness remain. The empirical findings reported in the following section provide a preliminary exploration of some of these issues. We used questionnaires completed by attorneys⁷⁶ who had participated in federal appellate civil mediation to examine whether their ratings of the mediators' skillfulness on seven different dimensions could differentiate among the mediators in terms of which mediators were more or less skillful on each dimension. In addition, we examined the extent to which these differences could be attributed to settlement or other mediation outcomes. Finally, we explored whether the attorneys' ratings were able to assess different aspects of mediator skillfulness. Thus, the focus of our inquiry was whether participant assessments might be a useful *method* for evaluating mediator performance, not on what the *content* of those assessments should be.

A. The Mediation Process

The Office of the Circuit Mediators for the U.S. Court of Appeals for the Sixth Circuit⁷⁷ is authorized to conduct mediation conferences to "consider the possibility of settlement, the simplification of the issues, and any other matters which . . . may aid in the handling of the disposition of the proceedings"⁷⁸ in order to reduce case management burdens and save judicial time.⁷⁹ Most of the civil cases eligible for mediation⁸⁰ are routinely scheduled for an initial mediation conference.⁸¹ Lead counsel for all parties are required to participate in the initial conference and are asked to come prepared to articulate their view of the merits of the case as well as their clients' settlement interests and needs.⁸² Clients are welcome but are not required to participate in the initial mediation conference, and they do so in fewer than ten percent of cases. Because the cases are distributed

76. Clients were not surveyed because few had directly observed the mediators' performance. See *infra* notes 82-83 and accompanying text.

77. For more details about the development, purposes, and operation of the Office, see JAMES B. EAGLIN, THE PRE-ARGUMENT CONFERENCE PROGRAM IN THE SIXTH CIRCUIT COURT OF APPEALS: AN EVALUATION 11-22 (1990); ROBERT J. NIEMIC, MEDIATION & CONFERENCE PROGRAMS IN THE FEDERAL COURTS OF APPEALS 52-57 (1997) (describing mediation programs in other circuits as well); Robert W. Rack, Jr., *Pre-Argument Conferences in the Sixth Circuit Court of Appeals*, 15 U. TOL. L. REV. 921, 921-25 (1984) [hereinafter Rack, *Pre-Argument*]. Prior to 1996, the name of the Office was the "Pre-Argument Conference Program" and the circuit mediators were called "conference attorneys."

78. 6TH CIR. R. 18(c)(2).

79. EAGLIN, *supra* note 77, at 3. The Program was based on the assumption that the appellate process affords fewer opportunities for lawyers to casually discuss settlement than exist at the trial level, making it all the more difficult for attorneys to initiate and sustain settlement negotiations. *Id.* at 12.

80. Cases involving a prisoner, one or more *pro se* parties, and unresolved jurisdictional problems generally are not considered eligible for mediation, and federal agency appeals such as Social Security, Tax Court, and NLRB cases are not routinely scheduled for mediation. See Rack, *Pre-Argument*, *supra* note 77, at 925, 926 (explaining why some case types are excluded).

81. In about twenty-five percent of the cases, mediation is requested by one or more of the parties. Requests for mediation are not disclosed to the court or to the other parties in the case and are nearly always granted.

82. The scheduling notice and accompanying materials contain expectations and suggestions for attorney preparation for the conference.

throughout a four-state area, the initial conference is conducted by telephone rather than in person in ninety to ninety-five percent of the cases.⁸³

At the initial conference, all counsel are “present” on the telephone conference call for an opening “joint session;” the conference eventually breaks into individual caucuses in a majority of cases. The conference begins with the mediator reviewing the confidentiality rules⁸⁴ and the purposes of the conference. After a usually brief discussion of procedural questions or issues, the attorneys explain their views of the primary issues to be raised on appeal. The mediator encourages the attorneys to pose questions and to engage in an objective analysis of the probabilities of outcomes of key issues rather than in arguments and accusations. The mediator invites and tries to facilitate a collaborative appraisal of the settlement value of the case while also working with each side to explore their interests and generate settlement options. Although there is some variation among the individual mediators, their orientation is primarily facilitative and elicitive rather than directive.⁸⁵

The initial conference typically lasts sixty to ninety minutes. However, when clients are present or counsel are prepared to negotiate all the way to a settlement, the conference can last several hours. Most of the cases that continue in mediation beyond the initial conference involve private conversations between the mediator and each party separately, although in some cases all parties are brought back together to negotiate further or finalize an agreement.⁸⁶ The mediator continues to be involved in settlement negotiations until the case settles or it is clear that the case cannot be settled.

*B. The Mediators*⁸⁷

The Office requires all mediators to be lawyers with good legal analytical skills⁸⁸ because the subject matter and context of most appellate mediation is heavily law based.⁸⁹ Each newly hired mediator participates in a forty hour interest-based negotiation training course. In addition, the new mediator participates

83. The telephonic technology accommodates up to nine parties on the line simultaneously and permits the mediator to put one or more parties on hold while caucusing with the other party or parties.

84. Communications between the parties during mediation, and between the mediator and the parties, are confidential.

85. See Leonard L. Riskin, *Who Decides What? Rethinking the Grid of Mediator Orientations*, 9 DISP. RESOL. MAG., Winter 2003, at 22, 24.

86. Fewer than five percent of the cases settle at the initial conference. About twenty to twenty-five percent of the cases end with or shortly after the initial conference, usually because all parties and the mediator agree that the case cannot be settled.

87. For additional details on the selection, training, and monitoring of the mediators, see Rack, *Thoughts*, *supra* note 63, at 614-25. The number of mediators in the Office has grown from one, at its inception, to three full-time and two part-time mediators currently.

88. There are no other specific educational, experience, or training requirements. Instead, the qualifications sought include life and work experience that demonstrate intelligence, integrity, initiative, maturity, and excellent interpersonal, collaborative, and communication skills.

89. The parties' BATNA (best alternative to a negotiated agreement) usually is their prospect on appeal. Arguments shaped for a panel of appellate judges typically turn on highly legalistic critiques of the trial court's procedures, evidentiary rulings, and interpretations of cases.

in an apprenticeship orientation that typically lasts a month or two.⁹⁰ Eventually, the mediator takes on cases of her or his own, with the Chief Circuit Mediator observing or co-mediating. The new mediator will continue to debrief with an experienced mediator for as long as he or she finds it useful.

Professional development for all mediators is ongoing. In addition to periodically attending formal outside mediation and substantive law programs, the mediators discuss cases at bi-weekly staff meetings. These meetings help standardize policies and procedures and provide an opportunity to share mediation strategies and to discuss cases and legal theories that bear on many of the Office's cases. The mediators also communicate with each other daily on an informal basis.

The Office relies on four information sources to monitor the performance of the program and its mediators and to provide the mediators with feedback they can use to review their own performance. The first source of information, reviewed by all of the mediators, is monthly and year-end statistics showing the number of cases settled by each mediator.⁹¹ The second source of information is questionnaires periodically sent to all attorneys in cases recently terminated by the mediation program.⁹² All of the mediators review all questionnaires and discuss any interesting responses or patterns. The third source of information about possible problems or exemplary performance is informal, anecdotal reports from lawyers and judges.⁹³ Finally, the Chief Circuit Mediator occasionally observes parts of mediations and commonly discusses the specifics of individual cases with each mediator, thereby staying familiar with his or her style and performance.

C. The Survey Procedure and Respondents

The Office distributed⁹⁴ a four-page questionnaire⁹⁵ to all attorneys⁹⁶ who had participated in mediation in cases that were closed⁹⁷ between mid-September 2000

90. During that time, the mediator prepares for and participates in mediation sessions with the experienced mediators, starts learning the rules and procedures of the court and the Office, and gains familiarity with the most common types of cases and applicable case law.

91. Settlement statistics are not, however, directly related to formal performance reviews of the mediators or to their salaries.

92. Approximately every five years, the Office surveys attorney participants to provide them the opportunity to anonymously express their opinions about what they do and do not like about the Office's practices, policies, and procedures. See Rack, *Thoughts, supra* note 63, at 622 (discussing examples of how attorney feedback has shaped the policies of the program as well as the practices of individual mediators). The questionnaires usually are not subjected to statistical analysis and are not intended to provide a formal evaluation of the program. For an exception, see the program evaluation by EAGLIN, *supra* note 77.

93. The Chief Circuit Mediator periodically speaks to bar groups and attends judicial conferences, and commonly receives informal feedback at these meetings. Formal complaints to the court about the Office are rare; fewer than half a dozen have been received in twenty years.

94. Questionnaires were mailed with a letter from the Circuit Executive and a self-addressed, stamped envelope for return to that office. This was done to emphasize that the court, and not just the Office, considered the attorneys' responses to be important and to assure them of anonymity so that they would be more likely to respond and to give candid ratings.

95. The present questionnaire was one of the periodic surveys the Office conducts, see *supra* note 92, and was not designed specifically for the purpose of conducting a formal statistical evaluation of mediator performance. This version of the questionnaire, however, was the first to include a set of questions asking participants about individual mediators' specific skills.

and February 2001. Four hundred and five attorneys returned completed questionnaires, for a response rate of sixty-one percent. Virtually all attorneys (ninety-six percent) had prior experience with civil mediation, either in this court or elsewhere.⁹⁸

During the time period covered by the present survey, the Office had five staff mediators who were randomly assigned to cases.⁹⁹ Each mediator was roughly equally represented across the attorneys' questionnaires, ranging from a low of sixty-five completed questionnaires (sixteen percent) to a high of ninety completed questionnaires (twenty-two percent). At the time of the survey, one of the mediators had been serving as a mediator in this Office for twenty years, one for fourteen years, two for seven to nine years, and one for two years. Because full-time mediators in the Office typically mediated more than 200 cases a year, most of the mediators had extensive experience mediating appellate cases at the time of the survey.

V. AN EMPIRICAL EXAMINATION OF ATTORNEYS' ASSESSMENTS OF MEDIATOR SKILLFULNESS

A. Attorneys' Ratings of the Mediators' Skillfulness

The attorneys were asked to rate the skillfulness of the mediator in their case on each of seven dimensions that the mediators had selected as being those which best represented the skills and objectives they tried to bring to their work.¹⁰⁰ These dimensions were: (1) reducing tensions or animosity between the participants, (2) helping participants objectively evaluate the strengths and weaknesses of their arguments and appraise the settlement value of the case, (3) identifying parties' underlying and/or unexpressed interests, motivations, and concerns, (4) uncovering previously unexpressed flexibility and willingness to compromise, (5) generating new ideas and options for settlement, (6) overcoming obstacles and impasses in the negotiations, and (7) guiding the negotiating process.

On each of these seven dimensions, the largest percentage of attorneys—ranging from thirty-nine percent to fifty-two percent, depending on the dimen-

96. Clients were not surveyed because they seldom had participated directly in mediation and, as a result, had not observed the mediator. *See supra* notes 82-83 and accompanying text.

97. Questionnaires typically were mailed to attorneys within two weeks after the mediators closed their files in a case. However, the amount of time that elapsed between when the mediators' work was completed and when the case was closed varied considerably, and tended to be shorter for cases that had not settled than for those that had settled. If the case had not settled, the mediators generally closed the case when they determined that further settlement efforts would not be productive. If the case had settled, the mediators generally waited to close the case until the settlement was executed and stipulations authorizing dismissal of the appeal were received, which could take from several weeks to several months.

98. Approximately one-third of the attorneys previously had been involved in mediation in fewer than ten cases, another third had mediated ten to twenty cases, and another third had mediated more than twenty cases.

99. Two of the mediators worked full-time; three worked part-time.

100. Mediation programs in different settings may consider other skill measures to be more appropriate for their context. *See supra* note 45 and accompanying text.

sion—rated the mediator as “very skillful.”¹⁰¹ From twenty-one percent to thirty-four percent of the attorneys rated the mediator as “extremely skillful.” Thus, a majority of attorneys rated the mediator as either “very” or “extremely” skillful on each of the dimensions. From fifteen percent to thirty-four percent of the attorneys rated the mediator as “somewhat skillful.” Few attorneys (from one percent to seven percent) rated the mediator as “not at all skillful” on any of the dimensions.

SKILL DIMENSIONS	not at all skillful	somewhat skillful	very skillful	extremely skillful	Mean (responses)
guiding negotiation process	3%	15%	48%	34%	3.14 (369)
reducing tensions, animosity	1%	18%	52%	28%	3.08 (303)
identifying interests, concerns	3%	22%	47%	28%	3.00 (351)
uncovering flexibility, compromise	6%	25%	42%	27%	2.90 (328)
helping parties evaluate case	4%	28%	47%	22%	2.87 (363)
generating new ideas, options	7%	32%	39%	22%	2.75 (310)
overcoming obstacles, impasses	6%	34%	39%	21%	2.74 (311)

Note: Rating scale ranged from 1 (not at all skillful) to 4 (extremely skillful).

TABLE 1. SKILLFULNESS RATINGS, ALL MEDIATORS

Attorneys' ratings on these seven mediator skill dimensions were highly intercorrelated.¹⁰² That is, attorneys who rated the mediator as highly skillful on one dimension also tended to rate her or him as highly skillful on the other dimen-

101. See *infra* Table 1. Given the extensive amount of experience most of these mediators had, see *supra* Part IV.C, relatively few ratings at the low end of skillfulness were expected. Accordingly, both “very” and “extremely” skillful were included as points on the rating scale to give the attorneys more opportunity to differentiate among the mediators at the high end of skillfulness. The attorneys also were given the response option of “unable to form any opinion,” and from 33 to 100 chose that option, depending on the skill dimension. Attorneys who said they were unable to form an opinion for a particular skill dimension were not included in any analyses involving that dimension.

102. *r*'s ranged from .508 to .773, *p*'s < .001. The Pearson *r* statistic assesses whether an apparent relationship between two variables is a “true” relationship (i.e., is statistically significant) or merely reflects chance variation. The conventional level of probability (*p*) for determining the statistical significance of findings is the .05 level (i.e., *p* < .05). Findings of *p* < .10 are considered “marginally” significant—the difference is not statistically significant, but is worth mentioning. The value of *r* (the correlation coefficient) indicates the strength of the relationship and ranges from +1.00 to -1.00, with 0.00 representing no relationship between the variables. See RICHARD P. RUNYON & AUDREY HABER, FUNDAMENTALS OF BEHAVIORAL STATISTICS 140-42, 229-31 (5th ed. 1984). As a rough guide to interpreting the size of correlation coefficients, *r* = .10 is considered a small relationship, *r* = .30 is considered medium, and *r* = .50 is considered a large relationship. See ROBERT ROSENTHAL & RALPH L. ROSNOW, ESSENTIALS OF BEHAVIORAL RESEARCH 361 (1984).

sions, and those who rated the mediator's skill as low on one dimension tended to rate his or her skill as low on the other dimensions.

B. Can the Ratings Differentiate among the Skill Dimensions?

The above pattern of high intercorrelations might suggest that attorneys' ratings were strongly influenced by their general, overall reaction to the mediator or the mediation process.¹⁰³ If so, then attorneys who had a favorable impression of the mediator or the process might tend to rate the mediator as highly skilled on each dimension, whereas those who were less pleased with the mediator or the process might tend to rate the mediator as less skilled on each dimension. This might be especially likely to happen if the attorneys made their skill ratings without careful consideration. The usefulness of the skill ratings for monitoring mediator performance would be limited if attorneys were not critically differentiating among the seven skills when rating the mediators.

A series of analyses demonstrated that the attorneys did not rate a given mediator as being uniformly skillful across all seven dimensions. Instead, for each mediator, attorneys' ratings across the skill dimensions were statistically significantly different.¹⁰⁴ Thus, attorneys rated individual mediators as being more skillful on some dimensions than others. For each mediator, the size of the differences in skillfulness ratings among the different dimensions was small to moderate.

Although which skills were statistically significantly different from which other skills varied across the mediators, the following common patterns emerged. The attorneys rated each mediator as being significantly more skillful at guiding the negotiation process, reducing tensions, and identifying interests¹⁰⁵ than at overcoming obstacles and generating new ideas.¹⁰⁶ The attorneys tended to give intermediate ratings to the mediator's skillfulness at uncovering flexibility and helping the parties evaluate the case.¹⁰⁷ This relatively similar pattern of skill

103. This pattern might also suggest that the outcome of the mediation influenced attorneys' responses. For the effect of settlement on the skill ratings, see *infra* notes 114-15 and accompanying text, and Part IV.D.

104. Mediator1, $F(6, 318) = 5.21, p < .001$; Mediator2, $F(6, 252) = 3.88, p < .01$; Mediator3, $F(6, 210) = 2.59, p < .05$; Mediator4, $F(6, 270) = 8.04, p < .001$; Mediator5, $F(6, 276) = 6.32, p < .001$. In this article, we refer to the mediators by code number to ensure their anonymity. The analysis of variance (which produces an F statistic) determines whether an apparent difference between individuals (e.g., among the five mediators) or between groups (e.g., between cases that settled and those that did not) is a "true" difference (i.e., is statistically significant) or merely reflects chance variation. See RUNYON & HABER, *supra* note 102, at 316-20. See *supra* Table 1, for the mean rating on each skill dimension for all mediators combined; see *infra* Table 2, for the mean rating on each skill dimension for each mediator.

105. For three of the mediators, there were no statistically significant differences among these three ratings. The other two mediators were rated as more skillful at guiding the negotiation process than at reducing tensions and identifying interests; ratings of the latter two skills did not differ significantly.

106. There were no statistically significant differences between ratings of the mediator's skillfulness at overcoming obstacles and at generating new ideas for any of the mediators.

107. There were no statistically significant differences, for any of the mediators, between ratings of the mediator's skillfulness at uncovering flexibility and at helping the parties evaluate the case. For all but one mediator, both of these skills were rated significantly lower than skillfulness at guiding the negotiation process. Whether attorneys' ratings of these skills differed from their rating of any of the other skills, and from which particular skills, varied across the mediators.

ratings for each of the mediators might be due to Mediation Office practices that maintain a relatively consistent approach across its mediators.¹⁰⁸

There are several possible explanations for the pattern of which particular skill dimensions tended to receive the highest (guiding the negotiation process, reducing tensions, identifying interests), intermediate (uncovering flexibility, helping parties evaluate the case), and lowest ratings (generating new ideas and settlement options, overcoming obstacles and impasses). Because the mediators' approach tended to focus on the process more than on specific outcomes,¹⁰⁹ perhaps the skills most consistent with this approach received more emphasis and refinement and, as a result, the mediators were more proficient on these dimensions.¹¹⁰ Or perhaps the context of appellate mediation played a role:¹¹¹ the high rating for reducing tensions and animosities might reflect that mediators had to put significant effort into that area, while the low rating for generating new ideas and settlement options might reflect that the mediators did not work as hard in this area due to their experience that doing so was rarely productive in these cases.¹¹² Or perhaps the pattern of skill ratings suggests that attorneys used different implicit standards when assessing different dimensions. For instance, if attorneys felt that mediators needed to be more skillful in certain areas (e.g., overcoming obstacles) than others (e.g., guiding the negotiation process) in order to resolve intractable appellate cases, a mediator who was in fact equally skillful on these dimensions might nonetheless receive lower ratings on those dimensions to which the attorneys applied a higher standard of performance. Alternatively, the pattern of skill ratings might reflect attorneys' expectations about what the mediators would do. For example, if the attorneys expected a more directive approach, they might have rated the mediators as less skilled on those actions that the mediators engaged in less intensely than expected.¹¹³

108. See *supra* Part IV.A., B.

109. See *supra* note 85 and accompanying text.

110. Or perhaps the mediators adopted a process-oriented approach *because* they were more skilled in these areas. See Honeyman, *Consensus*, *supra* note 23, at 301; Honeyman, *Common Core*, *supra* note 55, at 78.

111. Typically, neither the lawyers nor the parties approach appellate mediation with expectations of, or desire for, creative resolutions that would build on future relationships or other mutual interests. By the time of appeal, parties' positions usually have become highly polarized by years of contentious adversarial proceedings, and most have long ago concluded that their adversaries are unreasonable, intractable, and probably impossibly dishonest. As a consequence, most parties seem to just want to cash out and never have to deal with each other again.

112. It does not appear that the pattern of which skill dimensions received higher or lower ratings simply reflects the frequency with which the mediators performed certain actions (rather than how well they performed those actions). First, attorneys who indicated they were "unable to form any opinion" about the mediator's skillfulness on any dimension were not included in analyses involving that dimension. See *supra* note 101. Second, the number of attorneys who said they could not form an opinion about the mediators' skillfulness on the different dimensions did not directly track their ratings on those dimensions. For instance, the two dimensions that received the highest skillfulness ratings were those for which the fewest (33) and the most (100) attorneys, respectively, said they could not form an opinion. Third, while attorneys might conceivably give low skillfulness ratings to mediator actions that occurred infrequently, based on a presumption of lack of skill in that area, it does not necessarily follow that attorneys would assume competence and, thus, give high skillfulness ratings to those mediator actions that occurred frequently.

113. We did not have information on the style of mediation the attorneys previously had experienced or on their expectations regarding mediator style. The *extent* of the attorneys' prior mediation experience, however, was not related to their skillfulness ratings.

The differentiation among the skills of each mediator was smaller for attorneys whose case was settled¹¹⁴ than for attorneys whose case was not settled.¹¹⁵ We can only speculate why this might be so. Less differentiation among the skills by attorneys in settled cases could reflect a “halo effect” of more consistently favorable ratings that might accompany settlement. Or this pattern might indicate that these attorneys had less detailed memories of mediation, as more time typically had elapsed between mediation and receipt of the questionnaire in settled cases than in non-settled cases.¹¹⁶ Or perhaps these findings suggest that settlement was more likely when the mediator’s skill level across the different dimensions was more uniform than when it was more varied.

C. Can the Ratings Differentiate among the Mediators?

In order for attorneys’ ratings to be useful in monitoring mediator performance, the ratings would need to be able to differentiate among the different mediators. And indeed, the attorneys’ ratings of the mediators’ overall skillfulness—that is, their average level of skillfulness over all seven dimensions combined—varied significantly among the mediators.¹¹⁷ Mediator1 had a statistically significantly higher overall skillfulness rating than did Mediator3, Mediator4, and Mediator5,¹¹⁸ who did not differ significantly from each other. Mediator2 had a statistically significantly higher overall skillfulness rating than Mediator5,¹¹⁹ but did not differ significantly from any of the other mediators.

We next examined whether there were differences among the mediators in attorneys’ ratings of their skillfulness on each of the seven dimensions. The attorneys’ ratings of the mediators were statistically significantly different on five skill dimensions: (1) helping participants objectively evaluate the strengths and weaknesses of their arguments and appraise the settlement value of the case,¹²⁰ (2) overcoming obstacles and impasses in the negotiation,¹²¹ (3) reducing tensions or animosity between the participants,¹²² (4) identifying the parties’ underlying or unexpressed interests, motivations, and concerns,¹²³ and (5) uncovering previously unexpressed flexibility and willingness to compromise.¹²⁴ The attorneys’ ratings of the mediators were marginally significantly different¹²⁵ on the remaining two

114. The questionnaire did not distinguish between whether the case was settled or whether it was voluntarily dismissed. In this article, for the sake of brevity, we use “settled” to include both settlements and voluntary dismissals.

115. For three mediators, statistically significant differences in attorneys’ ratings across the skill dimensions were found both in cases that were not settled and those that were settled, but the differences were larger in the cases that were not settled. For the other two mediators, statistically significant differences in skill ratings were found only in cases that were not settled.

116. See *supra* note 97.

117. $F(4,221) = 3.43, p < .05, r = .243$. See *infra* Table 2, for the mean overall skillfulness rating for each mediator.

118. Respectively, $p = .051, p < .05, p < .01$.

119. $p < .05$.

120. $F(4,358) = 5.71, p < .01, r = .245$.

121. $F(4,306) = 4.44, p < .01, r = .234$.

122. $F(4,298) = 3.67, p < .01, r = .217$.

123. $F(4,346) = 3.17, p < .05, r = .188$.

124. $F(4,323) = 2.73, p < .05, r = .181$.

125. See *supra* note 102, for an explanation of “marginally” significant findings.

dimensions: (1) guiding the negotiation process¹²⁶ and (2) generating new ideas or options for settlement.¹²⁷ Overall, the size of the differences among the mediators was small to moderate.

Thus, the attorneys rated some mediators as being more skillful than others. We will briefly summarize the differences among the mediators across the seven skill dimensions.¹²⁸ Mediator1 received higher skillfulness ratings than Mediator3, Mediator4 and Mediator5 on six of the seven dimensions. The skillfulness ratings of Mediator1 and Mediator2 did not differ significantly on any dimension. Mediator2 tended to receive higher skillfulness ratings than one or more of the other Mediators (not including Mediator1), depending on the dimension. On most of the skill dimensions, attorneys' ratings of Mediator3, Mediator4, and Mediator5 did not differ significantly.

SKILL DIMENSIONS	M 1	M 2	M 3	M 4	M 5
guiding negotiation process	3.26	3.28	3.00	3.04	3.06
reducing tensions, animosity	3.29	3.14	2.92	3.10	2.87
identifying interests, concerns	3.25	3.05	2.90	2.97	2.83
uncovering flexibility, compromise	3.15	3.01	2.76	2.81	2.76
helping parties evaluate case	3.09	3.06	2.71	2.85	2.62
generating new ideas, options	2.93	2.90	2.64	2.60	2.64
overcoming obstacles, impasses	3.01	2.88	2.68	2.60	2.47
OVERALL SKILLFULNESS	3.19	3.04	2.91	2.87	2.73

Note: Rating scale ranged from 1 (not at all skillful) to 4 (extremely skillful).

TABLE 2. MEAN SKILLFULNESS RATINGS OF EACH MEDIATOR

D. Do the Ratings Simply Reflect Settlement or Other Mediation Outcomes?

When we examined the mediators' settlement rates, we found statistically significant differences among the mediators in the proportion of cases settled.¹²⁹ For attorneys' ratings of the mediators to be useful as a method for monitoring performance, the ratings would need to assess something other than simply whether or not the case was settled. Of course, more skillful mediators might have a higher settlement rate than less skillful mediators. But if the attorneys' ratings indicate nothing more than whether or not the case was settled, they would add no information about the mediators' performance beyond that which could more easily be obtained by tracking mediator settlement rates.

When we statistically controlled for whether or not the case was settled, statistically significant differences among the mediators in overall skillfulness re-

126. $F(4,364) = 2.19, p = .069, r = .153$.

127. $F(4,305) = 2.06, p = .086, r = .162$.

128. For the mean ratings for each mediator on each dimension, see *infra* Table 2. It is worth noting that, despite the differences among the mediators, more than half of the attorneys rated even the lowest-rated mediator as "very" or "extremely" skillful on six of the seven dimensions.

129. $F(4,400) = 3.10, p < .05, r = .173$.

mained,¹³⁰ and the specific pattern of differences among the mediators was virtually unchanged. Thus, differences among the mediators in overall skillfulness could not be attributed primarily to differences in the proportion of cases settled by each mediator.

For five of the seven individual skill dimensions, the patterns of differences among the mediators on each dimension remained when statistically controlling for whether or not the case was settled.¹³¹ Thus, differences among the mediators in attorneys' ratings of skillfulness at helping evaluate the case, overcoming obstacles, reducing tensions, identifying interests, and guiding the negotiation process were not due to differences in the mediators' settlement rates. For the other two dimensions, however, some of the variation across the mediators in skillfulness reflected differences in the proportion of cases settled by each mediator. When controlling for whether or not the case was settled, the differences among the mediators in attorneys' ratings of skillfulness at uncovering flexibility or willingness to compromise were only marginally significant,¹³² and there were no statistically significant differences among the mediators in attorneys' ratings of skillfulness at generating new ideas or options.¹³³ For the most part, however, differences among the mediators in skillfulness could not be attributed to differences in their settlement rates.

Settlement is not, of course, the only outcome that can be achieved in mediation. The attorneys also were asked whether the mediation procedure was helpful in several other respects. Sixty percent of attorneys said mediation was helpful in evaluating the appeal, and fifty-nine percent said it was helpful in clarifying issues in the appeal. Forty-nine percent of attorneys said mediation was helpful in dealing with any procedural problems, and twenty-nine percent said mediation was helpful in eliminating issues.¹³⁴ The attorneys' ratings of whether mediation was helpful in dealing with procedural problems¹³⁵ and in eliminating issues¹³⁶ varied significantly depending on which mediator had mediated their case, although their ratings of whether mediation was helpful in clarifying issues or in evaluating the appeal did not.

When we statistically controlled for whether or not the attorneys reported mediation was helpful in achieving each of these four outcomes, statistically significant or marginally significant differences among the mediators in attorneys' ratings of overall skillfulness remained.¹³⁷ Thus, differences among the mediators

130. $F(4,220) = 2.60, p < .05$.

131. Helping evaluate the case: $F(4,357) = 5.21, p < .001$; overcoming obstacles: $F(4,305) = 2.89, p < .05$; reducing tensions: $F(4,297) = 3.24, p < .05$; identifying interests: $F(4,345) = 2.96, p < .05$; guiding the negotiation process: $F(4,363) = 1.99, p = .096$.

132. $F(4,322) = 1.98, p = .097$. When settlement was not controlled for, these differences were statistically significant. See *supra* note 124 and accompanying text.

133. $F(4,304) = 1.78, p = .133$. When settlement was not controlled for, these differences were marginally significant. See *supra* note 127 and accompanying text.

134. It is worth noting that, of attorneys in cases that were not settled, fifty-one percent said mediation was helpful in clarifying issues, forty-six percent in evaluating the appeal, thirty-nine percent in dealing with procedural problems, and fifteen percent in eliminating issues.

135. $F(4,355) = 3.60, p < .01, r = .197$.

136. $F(4,361) = 3.35, p < .05, r = .190$.

137. When controlling for: eliminate issues ($F(4,203) = 2.27, p < .063$), clarify issues ($F(4,211) = 4.07, p < .01$), deal with procedural problems ($F(4,199) = 2.67, p < .05$), and evaluate the appeal ($F(4,212) = 3.15, p < .05$).

in overall skillfulness could not be attributed primarily to differences in the proportion of their cases in which the attorneys said mediation was helpful in evaluating the case, dealing with procedural problems, or clarifying or eliminating issues.

Taken together, the findings in this section suggest that attorneys' ratings of the mediator's skillfulness are tapping something other than simply whether or not their case settled or achieved certain other outcomes.

E. Are the Ratings Measuring "Skillfulness"?

Having found that the attorneys' ratings differentiated both among the seven dimensions and among the five mediators and did not simply mirror whether mediation led to settlement or to several other outcomes, we next examined whether ratings on these seven dimensions seemed to be measuring mediator "skillfulness." One would expect that greater mediator skillfulness would be strongly associated with a greater proportion of favorable mediation outcomes.¹³⁸ Accordingly, we examined the degree to which the set of seven ratings could distinguish between attorneys whose case settled or achieved other mediation outcomes and attorneys whose case did not settle or achieve those outcomes.¹³⁹

The set of seven mediator skillfulness ratings statistically significantly distinguished between attorneys whose case was settled and attorneys whose case was not settled, and did so with seventy percent accuracy.¹⁴⁰ These mediator skillfulness ratings also significantly distinguished between attorneys who said mediation was helpful and attorneys who said mediation was not helpful on each of the following four dimensions: eliminating issues (seventy-six percent accuracy),¹⁴¹ clarifying issues (seventy-five percent accuracy),¹⁴² evaluating the appeal (seventy-three percent accuracy),¹⁴³ and dealing with procedural problems (sixty-seven percent accuracy).¹⁴⁴ Thus, the seven dimensions appeared to capture me-

138. Tests of the validity of a measure (i.e., the degree to which differences in scores on the measure reflect true differences on the characteristic that we seek to measure) include examining whether the observed findings are consistent with predictions about the relationships between the measures of interest (here, the skill ratings) and other variables, including how accurately the measures predict meaningful criteria (e.g., mediation outcomes). See ROSENTHAL & ROSNOW, *supra* note 102, at 82; CLAIRE SELTZ ET AL., RESEARCH METHODS IN SOCIAL RELATIONS 169, 170, 173 (3d ed. 1976).

139. We were not able to examine the relationship between skillfulness ratings and perceived fairness. Because so few attorneys (12 of 405, or 3%) felt anything was unfair or inappropriate about the way in which mediation was conducted, analyses using that variable could produce potentially unreliable results. The comments of some of these attorneys were not related to the mediators' actions per se, but to the more general operation of the mediation program (e.g., the advisability of attempting mediation in the instant case; suggestions about the structure of mediation, such as to drop the opening statement or to have clients participate). The twelve attorneys who noted a concern were essentially evenly distributed across the five mediators.

140. $\chi^2(7) = 48.20, p < .001, r = .443$. Discriminant function analysis derives a linear combination of the predictor variables (here, the seven skillfulness ratings) that maximally distinguish between the groups (here, between attorneys in cases that settled and attorneys in cases that did not settle). The procedure holds the error rate constant over all variables in the analysis, thereby avoiding the inflation that would arise from multiple, separate analyses. See BARBARA G. TABACHNICK & LINDA S. FIDELL, USING MULTIVARIATE STATISTICS 292, 295-98 (1983).

141. $\chi^2(7) = 39.09, p < .001, r = .418$.

142. $\chi^2(7) = 55.36, p < .001, r = .480$.

143. $\chi^2(7) = 54.79, p < .001, r = .477$.

144. $\chi^2(7) = 35.83, p < .001, r = .405$.

diator skills that were highly related to achieving favorable outcomes in mediation.¹⁴⁵

We wanted to explore further whether the seven different dimensions seemed to capture different aspects of mediator skillfulness and whether each dimension seemed to measure the type of skill its wording would suggest it was measuring.¹⁴⁶ If the seven dimensions measure different types of mediator performance, one would expect that (a) they would contribute in different degrees to distinguishing between attorneys whose case achieved each of several mediation outcomes and attorneys whose case did not achieve each of those outcomes, and (b) the patterns of their relative contributions would differ across the different outcome measures in ways that would be logically consistent with the concepts each dimension is thought to represent.¹⁴⁷

First, we examined the pattern of contributions that the seven skill dimensions made to distinguishing between attorneys whose case was settled and attorneys whose case was not settled. The mediator skills that made by far the largest contribution to distinguishing between attorneys whose case was settled and those whose case was not settled were skillfulness at uncovering flexibility or willingness to compromise and at overcoming obstacles.¹⁴⁸ The mediator's skillfulness at guiding the negotiation process, helping the parties evaluate the appeal, and generating ideas and options also made statistically significant but smaller contributions to distinguishing between attorneys whose case was settled and those whose case was not settled. The mediator's skillfulness at reducing tensions and identifying interests did not contribute significantly to distinguishing between attorneys whose case settled and those whose case was not settled.

Next, we examined the patterns of contributions of the seven skill dimensions to distinguishing between attorneys who said mediation was helpful in eliminating issues, clarifying issues, dealing with procedural problems, or evaluating the appeal and attorneys who said mediation was not helpful in achieving these outcomes. The two skills that made the largest contribution to distinguishing between attorneys who said mediation was helpful in eliminating issues and those who said it was not were the mediator's skillfulness at uncovering flexibility and at overcoming obstacles.¹⁴⁹ By contrast, the two skills that made the largest con-

145. It might have been possible to obtain greater predictive accuracy of mediation outcomes by using a different set of skill measures. Nonetheless, the observed degree of discrimination between whether or not these outcomes were achieved is high, especially considering the inevitable variation among cases, attorneys, and mediators.

146. Some findings discussed earlier suggested that the skill dimensions measured different types of skillfulness. See *supra* notes 104-12 and accompanying text.

147. Another test of the validity of a measure is to examine whether there is divergence among measures of related but conceptually distinct constructs (also known as discriminant validity). See ROSENTHAL & ROSNOW, *supra* note 102, at 78-79; SELTZ ET AL., *supra* note 138, at 175.

148. See *infra* Table 3, for the canonical loadings. The canonical loadings (also called "structure coefficients") produced by discriminant function analysis indicate which variables make independent contributions to the overall discrimination as well as the relative strength of their contributions. The values of the loadings range from +1.00 to -1.00; by convention, .30 is the lowest cut-off for including variables to be used in interpreting the discriminant function. See WILLIAM R. KLECKA, *DISCRIMINANT ANALYSIS* 33 (1980); TABACHNICK & FIDELL, *supra* note 140, at 321.

149. See *infra* Table 3, for the canonical loadings. The mediator's skillfulness at generating new ideas, helping parties evaluate the case, and reducing tensions also made large contributions to distinguishing between attorneys who said mediation was helpful in eliminating issues and attorneys who

tribution to distinguishing between attorneys who said mediation was helpful in clarifying issues and those who said it was not were the mediator's skillfulness at generating new ideas and options and at helping the parties evaluate the case.¹⁵⁰ The three skills that made the largest contribution to distinguishing between attorneys who said mediation was helpful in evaluating the appeal and those who said it was not were the mediator's skillfulness at uncovering flexibility, helping the parties evaluate the case, and overcoming obstacles.¹⁵¹ The same three skills, although in a different order, made the largest contribution to distinguishing between attorneys who said mediation was helpful in dealing with procedural issues and those who said it was not.¹⁵²

SKILL DIMENSIONS	Case Settled	Eliminated Issues	Evaluated Appeal	Dealt with Procedure	Clarified Issues
uncover flexibility, compromise	.769	.879	.887	.865	.662
overcoming obstacles, impasses	.708	.875	.834	.786	.710
guiding negotiation process	.458	.567	.684	.627	.583
helping parties evaluate case	.380	.728	.867	.873	.832
generating new ideas, options	.365	.754	.752	.669	.954
reducing tensions, animosity	.274	.709	.550	.644	.527
identifying interests, concerns	.188	.525	.704	.610	.698
<i>Note:</i> Canonical loadings indicate the relative strength of the contributions that the skill ratings made to distinguishing between whether or not each outcome was achieved. ¹⁵³					

TABLE 3. CANONICAL LOADINGS

said it was not. Skillfulness at guiding negotiations and identifying interests made significant but smaller contributions.

150. See *infra* Table 3, for the canonical loadings. The mediator's skillfulness at overcoming obstacles, identifying interests, and uncovering flexibility also made large contributions to distinguishing between attorneys who said mediation was helpful in clarifying issues and attorneys who said it was not. Skillfulness at guiding negotiations and reducing tensions made significant but smaller contributions.

151. See *infra* Table 3, for the canonical loadings. The mediator's skillfulness at generating new ideas, identifying interests, and guiding negotiations also made large contributions to distinguishing between attorneys who said mediation was helpful in evaluating the appeal and attorneys who said it was not. Skillfulness at reducing tensions made a significant but smaller contribution.

152. See *infra* Table 3, for the canonical loadings. The mediator's skillfulness at generating new ideas, reducing tensions, guiding negotiations, and identifying interests also made large contributions to distinguishing between attorneys who said mediation was helpful in dealing with procedural issues and attorneys who said it was not.

153. See *supra* note 148.

Thus, as described above, the seven different mediator skill dimensions made different relative contributions to distinguishing between whether or not each of the mediation outcomes was achieved. In addition, the strength of each dimension's contribution varied across the different outcomes. For example, the two skills that made the largest contribution to whether mediation led to settlement made only intermediate contributions to whether mediation helped to clarify issues, and vice versa.¹⁵⁴ And skills that made little or no contribution to whether or not the case settled made sizeable contributions to the other mediation outcomes. Thus, the seven dimensions seem to be measuring different aspects of mediator performance.

For the most part, the patterns of contributions were consistent with what would be expected if the skill dimensions were measuring the type of skill that the question wording would suggest. For instance, uncovering willingness to compromise and overcoming obstacles seem to be skills that would be essential for settlement. The other mediator skills, while perhaps necessary in order to create conditions conducive to settlement, might not in and of themselves be sufficient to lead to settlement. As would be expected, skillfulness at helping parties evaluate the case made a large contribution to whether mediation was helpful in evaluating the appeal. Skillfulness at generating new ideas and helping to evaluate the case would logically seem to be relevant to helping clarify issues. This is not to suggest, however, that all of the patterns were what one might predict. For instance, one might have expected skillfulness at uncovering flexibility to be less highly related to helping to evaluate the appeal than it was, and skillfulness at overcoming obstacles to be less highly related to clarifying issues than it was.

VI. CONCLUSION: THE POTENTIAL USEFULNESS OF PARTICIPANT ASSESSMENTS

For participants' assessments to be useful in measuring mediator performance, their skillfulness ratings would need to be sufficiently discriminating to reveal differences among the different mediators and among the different skill dimensions. In addition, the assessments would need to be largely independent of whether or not the case settled. The findings of the present study showed that attorneys' assessments of mediators' skillfulness met these criteria.

Attorneys' ratings of the mediators' skillfulness on each of seven dimensions differentiated among the mediators—that is, the attorneys rated some mediators as more skillful than others. Importantly, the skillfulness ratings did not simply mirror whether favorable outcomes were obtained in mediation. For the most part, differences among the mediators in skillfulness ratings remained when statistically controlling for whether or not settlement or other outcomes were achieved. Thus, attorneys' assessments of the mediators appeared to be responsive to their skillfulness and were not overwhelmed by the outcome of mediation.

In addition, the attorneys rated each mediator as being more skillful on some dimensions than on others, suggesting that different mediators had somewhat different strengths. Thus, attorneys' assessments of the mediators appeared to be nuanced rather than simply reflecting their overall impression of the mediator.

154. See *supra* Table 3, for the canonical loadings.

Additional findings that the seven skill dimensions were differentially related to whether or not each of several mediation outcomes were achieved provided further evidence that the seven skill dimensions assessed different aspects of mediator skillfulness.

A question that remains is how participants' ratings would compare to other ratings of mediator performance, such as mediators' self-ratings or program administrators' ratings. Because it was not the original purpose of the attorney questionnaire to address this issue, mediator self-ratings and the program administrator's ratings were not obtained until substantially later.¹⁵⁵ The roughly two and one-half year gap between when the different sets of ratings were made raises serious issues of non-comparability, so we mention the comparisons here as merely suggestive. When we examine the rank ordering of the mediators on overall skillfulness (i.e., their average level of skillfulness over all seven dimensions combined, ranked from most skillful to least skillful), we find a similar pattern for the attorneys and the program administrator,¹⁵⁶ but a considerably different pattern based on the mediators' self-ratings.¹⁵⁷ These comparisons illustrate the potential limitations of relying on mediator self-assessments¹⁵⁸ and suggest that the attorneys' skillfulness ratings are at least not out of line with the program administrator's ratings. Future studies should obtain ratings from these groups at the same point in time so that more systematic and meaningful comparisons can be made.¹⁵⁹

In sum, the present findings suggest that attorneys' assessments of mediators' strengths and weaknesses provide information that mediators can use to enhance specific skills and that program administrators can use to monitor mediator performance.¹⁶⁰ The mediators and the program administrator in the present study found the use of participant assessments to be helpful and revealing in some additional ways. First, the process of identifying the skills they wanted to measure

155. In July 2003, the mediators were asked to rate their then-current skill levels on the seven dimensions, using the same rating scale the attorneys had used. The program administrator also rated the skillfulness of the mediators at that time.

156. The attorneys' and the administrator's rank orderings were the same for all mediators except one, the mediator most recently hired at the time of the attorney survey. It would not be surprising if, during the intervening years, her or his skills had improved proportionately more than those of the other mediators who had been with the Office longer.

157. One of the mediators appeared to under-rate his or her skills on most dimensions, while one or two of the mediators, admitting to not being modest, appeared to over-rate their skills on several dimensions.

158. See *supra* note 40. Several studies have found that mediators' reports of what they did during mediation or what was accomplished differed from those of attorney and client participants. See, e.g., Brazil, *supra* note 34, at 147; Wissler, *supra* note 24, at 656 n.55.

159. Other programs might want to have the mediators rate their skillfulness in each case. When aggregated across cases, this might provide a more accurate self-rating than a single rating made in the abstract. This approach would also permit direct comparisons of the mediator self-ratings and the attorneys' ratings in the same case.

160. Depending on their goals, needs, and resources, individual programs could use participant ratings in different ways. For instance, program administrators who want to know whether any mediator is performing at a level the program deems unacceptable (or exceptionally high) can simply look at participants' ratings of each mediator, aggregated across a sample of cases. Administrators or mediators who are interested in detecting more specific training needs would want to look in more detail at differences across skill areas or among mediators. For participants' ratings to be well-received by the mediators, it might be helpful for program administrators to create an atmosphere that emphasizes ongoing skill enhancement for all mediators and an approach that treats the assessments as constructive feedback rather than negative evaluation.

involved the whole office in a refreshing analysis of their purposes and objectives. Then, the process of reconciling the data with their preconceptions highlighted several possibly erroneous assumptions they were making and stimulated rethinking about the efficacy of certain mediation practices in this appellate context.

It is unclear, however, whether the present findings will generalize beyond this particular study to programs that use different skill ratings or that have different procedural, mediator, case, or participant characteristics. Below, we discuss several ways in which the present study might differ from other studies that could lead to different conclusions about the usefulness of participant ratings for monitoring mediator performance.

Participant assessments are likely to be less reliable for use in monitoring mediator performance if mediators are assigned to cases based on their skills rather than at random, as they were in the present study. For instance, a highly skilled mediator assigned to a complex case might receive skillfulness ratings that do not differ from those received by a less skilled mediator assigned to a less difficult case. Or the reverse could occur—mediators in tough cases might receive higher ratings because they had to exercise, or had more opportunity to demonstrate to participants, a higher level of skill. Accordingly, the use of participant ratings to compare the skillfulness of different mediators could be misleading if some mediators are routinely assigned to more (or less) difficult cases instead of if all mediators have a more comparable mix of cases.

The particular skills assessed in the present study were selected by the mediators as those specifically relevant for this program; other programs may want to assess different skill dimensions.¹⁶¹ The specific content or nature of the skills to be assessed might produce greater or lesser differentiation among mediators or among skills than observed in the present study. And the phrasing of the questions (or the labels on the rating scales) could differ in a number of ways, such as in their level of specificity or mediator versus process focus, which might produce different findings. For example, we might expect greater differentiation among mediators and among skill dimensions if participants rate whether or not the mediator performed specific actions (e.g., whether the mediator listened to both sides) than if participants rate more general aspects of mediator performance (e.g., whether the mediator was impartial) or of the mediation process (e.g., whether the process was fair).

The mediation program in the present study had a greater homogeneity of mediator orientation, practices, and procedures¹⁶² than is likely to exist in other mediation programs, such as those with a large roster of volunteer mediators.¹⁶³ If a particular action is rated differently by different participants, programs with greater internal variation might be less likely to find statistically significant differences in skillfulness ratings among the mediators. That is, if some participants

161. See *supra* notes 40, 45 and accompanying text. Although this program is typical of appellate mediation programs, the nature of mediation in appellate cases is substantially different from that in other court contexts, such as in small claims and general jurisdiction cases. See generally Roselle L. Wissler, *The Effectiveness of Court-Connected Dispute Resolution in Civil Cases*, CONFLICT RESOL. Q. (forthcoming 2004).

162. See *supra* PART IV.A., B.

163. See, e.g., Brazil, *supra* note 8, at 801-02 (discussing how ADR service-delivery models can differ in uniformity).

rate a certain behavior as being highly skillful while other participants give low ratings to the same behavior, the overall assessment of each mediator might not yield a meaningful pattern, resulting in less differentiation among the mediators.¹⁶⁴ If, on the other hand, most participants consistently rated one approach as being highly skillful and another approach as being less skillful, the variability in the ratings of each mediator would be reduced while differences between mediators who use different approaches would be increased, resulting in greater differentiation among mediators. Thus, whether participant assessments will be more or less useful in programs where mediators use a greater variety of procedures and approaches remains to be seen.

One factor thought to affect the usefulness of participant assessments is whether participants have enough knowledge about what they should expect from a competent mediator to make meaningful ratings.¹⁶⁵ The participants in the present study had a substantial amount of prior experience with mediation—approximately two-thirds had participated in civil mediation in ten or more cases. The extent of the attorneys' prior mediation experience, however, was not related to their skillfulness ratings. Future studies need to examine whether the ratings of attorneys or clients who have little or no mediation experience would also successfully differentiate among mediators and among skill dimensions.¹⁶⁶

Another factor that might affect the usefulness of participants' ratings is whether the participants take the time and effort to give thoughtful responses and make critical distinctions.¹⁶⁷ In the present study, the questionnaires were mailed to the attorneys several weeks to several months after the mediation session for them to complete on their own timetable. The present findings suggest that longer time intervals might reduce the degree of differentiation in participants' ratings.¹⁶⁸ It remains to be seen how participants' ratings would be affected if they instead were obtained as part of an exit survey to be completed immediately after mediation, when participants might have limited time available for that task and when they would have had less opportunity to reflect on the mediator's performance.¹⁶⁹

Thus, as the above discussion demonstrates, the usefulness of participant questionnaires needs to be empirically and systematically examined in a variety of different contexts in order to determine the conditions under which they can be most useful as a means for measuring and monitoring mediator performance. The present findings are encouraging and suggest that individual mediators as well as

164. By definition, finding statistically significant differences among the mediators indicates that the variability in participants' ratings between mediators is greater than the variability in the ratings of each mediator. See RUNYON & HABER, *supra* note 102, at 316.

165. See Honeyman, *Evaluating, supra* note 8, at 25, 33.

166. An additional question is whether the skills that mediators view as most important are those that matter most to participants. See McEwen, *supra* note 23, at 319.

167. See *supra* note 71 and accompanying text. And, of course, participants' ratings will be useful only if those who respond are representative of all participants and give candid responses. Participants will be more likely to complete a questionnaire and to provide candid assessments if they are assured that their responses are anonymous and that honest assessments are both wanted and important to the court and the mediation program. See FLOYD J. FOWLER, JR., *SURVEY RESEARCH METHODS* 52 (1988); SELTZ ET AL., *supra* note 138, at 165-66.

168. See *supra* note 116 and accompanying text.

169. Obviously, questionnaires completed after an initial session and before the completion of the mediation process would provide an incomplete and potentially misleading assessment of the mediator.

2004]

Assessing Mediator Performance

257

mediation program administrators can gain valuable information from participant assessments.

