External Evaluation of the National Institute on Disability and Rehabilitation Research and Its Grantees Letter Report

ISBN
978-0-309-21715-6

40 pages
8 1/2 x 11
2011

Committee on the External Evaluation of NIDRR and Its Grantees; National Research Council

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Division of Behavioral and Social Sciences and Education
Board on Human-Systems Integration

500 Fifth Street, NW
Washington, DC 20001
Phone: 202 334 2678
Fax: 202 334 2210
Email: bohsi@nas.edu
www.nationalacademies.org

July 8, 2011

Ms. Mary Darnell
Contracting Officer's Representative
U.S. Department of Education
Office of Special Education and Rehabilitative Services
550 12th Street, SW
Washington, DC 20202

Dear Ms. Darnell:

At the request of the National Institute on Disability and Rehabilitation Research (NIDRR) within the Office of Special Education and Rehabilitative Services, U.S. Department of Education, the Board on Human-Systems Integration of the National Research Council (NRC) convened an ad hoc committee to conduct an evaluation of aspects of NIDRR's program. Specifically, the Committee on the External Evaluation of NIDRR and Its Grantees was charged to review NIDRR's priority-setting, peer review, and grant management processes, develop an overall framework and evaluation design for the review of grantee outputs for a sample of 30 grantees, conduct a review of the sampled grantee outputs, and assess the output review process. (For a list of committee members, see Attachment A.)

The results of this project will be presented in a final report that will include a description and assessment of NIDRR's priority-setting, peer review, and grant management processes and the quality of the grantees' outputs. The committee's evaluation is nearing completion, and the committee plans to deliver the final report in fall 2011. However, knowing that NIDRR plans to move forward with an additional evaluation cycle prior to the delivery of the committee's final report, we wanted to provide the agency with information that could inform future evaluation design. This letter report is therefore limited in scope to discussing the procedures the committee used in its output evaluation, its assessment of those procedures, and recommendations for future evaluations. (See Attachment B for the names of the reviewers of this letter report.)

BACKGROUND ON NIDRR AND COMMITTEE CHARGE

The National Institute on Disability and Rehabilitation Research is the principal federal agency that funds applied research and development to improve the lives and functioning of persons with disabilities (Office of Special Education and Rehabilitative Services, 2007). NIDRR was established by the 1978 amendments to the Rehabilitation Act of 1973 and is one of three components of the Office of Special Education and Rehabilitative Services at the U.S. Department of Education (Office of Special Education and Rehabilitative Services, 2007).

NIDRR has conducted various efforts to assess its portfolio and grant results and hold its programs accountable for results. In 2009, NIDRR requested that the NRC conduct an evaluation of NIDRR and its grantees. The charge to the committee included two major components. The

first component, termed the "process evaluation," involved examining NIDRR's priority-writing process, its practices for the peer review of grant applications, and grant management processes. This component will be discussed in the committee's final report and is not covered in this letter report.

The second component, termed the "summative evaluation," involved the assessment of grantee outputs. The key question of the summative evaluation was articulated by NIDRR as follows: To what extent are the final outputs from NIDRR grants of high quality? The major portion of this component will similarly be covered in the committee's final report and is not covered in this letter report. However, one element in the summative evaluation involved a committee self-assessment of the methods it developed to conduct the summative evaluation and the identification of implications for future reviews. This element is the sole focus of this letter report, which is organized into three main sections. The first section summarizes the methods and procedures the committee used in the summative evaluation. The second section discusses the committee's assessment of these methods. The third section offers recommendations for future evaluations of NIDRR and its grantees.

SUMMARY OF METHODS DEVELOPED FOR ASSESSING THE QUALITY OF OUTPUTS

As noted above, the summative evaluation component involved an assessment of a wide range of grantee outputs, which are defined and categorized by NIDRR as follows.

1. Publications (e.g., research reports and other publications in peer-reviewed and nonpeer-reviewed publications).
2. Tools, measures, and intervention protocols (e.g., instruments or processes created to acquire quantitative or qualitative information, knowledge, or data on a specific disability or rehabilitation issue; or to provide a rehabilitative intervention).
3. Technology products and devices (e.g., industry standards/guidelines, software/netware, inventions, patents/licenses/patent disclosures, working prototypes, product(s) evaluated or field-tested, product(s) transferred to industry for potential commercialization, product(s) in the marketplace).
4. Informational products (e.g., training manuals or curricula, fact sheets, newsletters, audiovisual materials, marketing tools, educational aids, websites or other Internet sites that were produced in conjunction with research and development, training, dissemination, knowledge translation, and/or consumer involvement activities).

Committee members reviewed 148 outputs: 103 publications (category 1); 9 tools, measures, and intervention protocols (category 2); 9 technology products and devices (category 3), and 27 information products (category 4).

To prepare for the output review, the committee first developed a set of criteria and dimensions under those criteria that would be used to assess the quality of outputs. Second, the committee developed a questionnaire to assist grantees in nominating outputs to be reviewed and to give them the opportunity to provide supplemental descriptive information about each of the nominated outputs, along with the outputs themselves. Third, a sampling plan was developed to select grantees who would be invited to participate in the evaluation. Fourth, the committee staff worked with grantees who agreed to participate to gather and catalogue the outputs and supplemental information that were submitted for the committee's review. Fifth, the committee assessed the outputs through an expert review process that was based on direct review of the

outputs and any supplemental information provided by the grantees. These five steps in the study process are each described in detail below.

Quality Criteria Development

A key element of the summative evaluation was to respond to NIDRR's request to develop criteria for assessing the quality of its grantees' outputs.[1] In developing the criteria, the committee drew on its own research expertise, recommendations of the external advisory group convened by NIDRR while planning this NRC evaluation (National Institute on Disability and Rehabilitation Research, 2008), and methods used in other NRC and international studies that have evaluated federal research programs (see, e.g., Bernstein et al., 2007; Canadian Academy of Health Sciences, 2009; Chien, Chen, and Chen, 2009; Ismail, Tiessen, and Wooding, 2010; National Research Council, 2007; Wooding and Starkey, 2010; Wooding et al., 2009). The committee developed four criteria:

1. *Technical quality of output* The technical quality of outputs was assessed using dimensions that included applying standards of science and technology, appropriate methodology (quantitative or qualitative design and analysis), and degree of accessibility and usability.
2. *Advancement of knowledge or the field* The dimensions used to assess advancement of the knowledge base or of the field (e.g., research, practice, or policy as relevant) included scientific advancement of methods, tools, and theory; developing new information or technologies; closing an identified gap; and using methods and approaches that were innovative or novel.
3. *Likely or demonstrated impact* This criterion was used to assess the likely or demonstrated impact of outputs on science (journal impact, citations); consumers (for people with disabilities: health, quality of life, participation), provider practice, health and social systems, social and health policy, or the private sector or commercialization.
4. *Dissemination* The dimensions of dissemination included the identification and tailoring of materials for reaching different audience/user types; collaboration with audience/users in identifying content and medium needs/preferences; delivery of information through multiple media types and sources for optimal reach and accessibility; evaluation of dissemination efforts and impacts; and commercialization/patenting of devices, if applicable.

A 7-point scale was used to rate the criteria at varying levels of quality: 1 indicated poor quality, 4 indicated good quality, and 7 indicated excellent quality. A rating of 4 meant that the output solidly fell in the range of meeting expectations for good quality. See Box 1 for examples of quality indicators considered by committee members in determining each criterion score. These examples are not intended to be exhaustive, but to illustrate the attributes of outputs that were considered in the committee's review. In rating the outputs, committee members drew from their scientific expertise to consider the output's quality with respect to the dimensions under each criterion.

[1]The development of the criteria was informed by open session discussions in which NIDRR staff were present.

BOX 1
Examples of Quality Indicators Considered in Determining Output Scores

Technical Quality
- Strength of literature review and framing of issues
- Competence of design, considering the research question and other parameters of the study
- Quality of measurement planning and description
- Analytic methods and interpretation; degree to which recommendations for change were drawn clearly from the analysis
- Description of feasibility, usability, accessibility, and consumer satisfaction testing

Advancement of Knowledge/Practice
- Degree to which a ground-breaking and innovative approach is presented
- Application of a formal test of a hypothesis regarding a technique used widely in the field to improve practice
- Level of advancement and improvement to current classification systems
- Usefulness of descriptive base of information about factors associated with a condition
- Novelty of ways of studying a condition that can be applied to developing new models, training, or research

Likely or Demonstrated Impact
- Degree to which output is well cited or has promise to be (for newer articles)
- Potential to improve the lives of persons with disabilities through increasing accessibility
- Possibly transformative clinical and policy implications
- Potential for building capacity, lowering costs, commercialization, etc.
- Influence on the direction of research, use in the field, or capacity of the field

Dissemination
- Method and scope of dissemination
- Description of the evidence of dissemination (e.g., numbers distributed to different audiences);
- Level of strategic dissemination to target audiences when needed
- Evidence of reaching the target audience
- Degree to which appropriate multiple media outlets were used, such as webinars, TV coverage, senate testimony, website, DVD, and/or social network sites.

Grantee Questionnaire

NIDRR supplied the committee with information gathered from grantees in their Annual Performance Reports (APRs) and final reports (Research Triangle International, 2009). Grantees are required to complete APRs annually to report on their progress. At the end of a grant, they must complete a final report. To supplement the APRs and final reports provided by NIDRR, the committee developed a grantee questionnaire (see Attachment C). The first part of the questionnaire asked grantees to list each of the projects under the grant and nominate the top two outputs from each project that reflected the grants' best achievements. The questionnaire specified that outputs were to be drawn from the four categories defined in the APR (Research Triangle Institute, 2009), described above: (1) publications; (2) tools, measures, and intervention protocols; (3) technology products and devices; and (4) informational products.

The questionnaire instructions indicated that the committee would prefer to review one publication and one other type of output for each project under a grant, but that grantees could submit two publications for review if that was the only type of output for a project. The questionnaire asked the grantees to submit the actual outputs for the committee's review. If the output was a website, a tool, or a technology device that had to be demonstrated, grantees were

asked to provide descriptive information, pictures, or links to websites for the committee's direct review.

The second part of the questionnaire included a series of questions to elicit more in-depth descriptions of the outputs if needed and to provide supplemental evidence of the output's technical quality, how it advanced knowledge or practice, its likely or demonstrated impact, and about how it was disseminated. This type of information needed for a comprehensive assessment of the output would not always be apparent in reviewing the output in isolation.

For supplemental information on *technical quality*, grantees were asked to describe examples, such as the approach or method used in its development; relevant peer recognition; receipt of a patent, approval by the Food and Drug Administration, or use of the output in standards development; and evidence of the usability and accessibility of the output. For supplemental information on *advancement of knowledge or the field*, grantees were asked to discuss the importance of the original question or issue and describe how the output had advanced knowledge in such arenas as making discoveries; providing new information; establishing theories, measures, and methods; closing gaps in the knowledge base; and developing new interventions, products, technology, and environmental adaptations. For supplemental information on *likely or demonstrated impact*, grantees were instructed to describe the output's potential or actual impact on science, people with disabilities, provider practice, health and social systems, social and health policy, private sector/commercialization, capacity building, and any other relevant arenas. For supplemental information about *dissemination*, grantees were asked to describe the stage and scope (e.g., local, regional, national) of dissemination efforts, specific dissemination activities, any identification and tailoring of materials for particular audiences, efforts to collaborate with particular audiences or user communities to identify content and medium needs and preferences, and the delivery of information through multiple media types. Grantees were also asked to provide information from evaluations they may have conducted of their dissemination efforts and impacts (e.g., results of audience feedback or satisfaction surveys).

The committee piloted the questionnaire on one NIDRR grant that had ended in 2008 and was outside the sampling pool (described below). Operating through subgroups, the committee assessed five outputs of this grant, which consisted of publications, an assessment package, a working prototype, and a fact sheet. As a result of that assessment, the committee revised the questionnaire by collapsing some of the dimensions from an original six criteria into the four final criteria.[2]

To further supplement the grantee questionnaire in assessing the likely impact of published articles, the committee also used such sources as Scopus and the Web of Science to determine the journal impact factor and the number of citations of a particular article.

Sampling

NIDRR provided the committee with a data set of grantee information that consisted of all grants ending in years 2006−2010 (n = 248). Table 1 shows, within each program mechanism: the number of grants and the corresponding proportion of all NIDRR grants, the mean duration of grants, and the total funds expended and proportion of all funds expended. The last five

[2]An original criterion on output usability was collapsed into the final technical quality criterion. Another original criterion on consumer and audience involvement was restructured as dimensions of the other criteria. For example, the technical quality criterion now includes a dimension on "evidence of usability and accessibility." The impact criterion includes a dimension on "impact on people with disabilities." The dissemination criterion includes a dimension on "tailoring materials to audiences" and "collaboration with users."

columns of the table show the number of grants that ended in each year from 2006 to 2010. Highlighted in these last five columns is a subset of 111 grants that comprised the sampling pool from which 30 grants were randomly sampled for the summative evaluation.

The committee used the smaller subset of all NIDRR grants as the sampling pool because of its charge and preliminary analysis of the data. The committee was directed by its charge to draw a sample of 30 grants ending in 2009 that reflected the range of work conducted across NIDRR's 14 program mechanisms. However, as can be seen in Table 1, several program mechanisms did not have at least two grants ending in 2009: the three Model Systems (MS) mechanisms, Disability Business Technical Assistance Centers (DBTAC), Disability and Rehabilitation Research Projects-Knowledge Translation (DRRP-KT), Advanced Rehabilitation Research Training (ARRT), and grants under the Disability and Rehabilitation Research Projects Program (DRRP)-Section 21.

Because the MS grant mechanisms support some of NIDRR's flagship programs, including traumatic brain injury (MS-TBI), spinal cord injury (MS-SCI), and burn injuries (MS-Burn), adjustments were made to the sampling pool to ensure that these programs would be included in the sample. The committee thus went back to the nearest year that yielded a total of at least two grants, which was 2008 for MS-Burn and MS-TBI (n = 5 for Burn; n = 9 for MS-TBI, with 1 in 2009 and 8 in 2008) and 2007 for MS-SCI (n = 9) and included these grants in the pool. The DBTAC, DRRP-KT, ARRT, and DRRP-Section 21 were excluded from the pool for this first cycle of evaluations. Small Business Innovation Research, Phase I, grants were also excluded from the sampling pool because they do not produce "outputs" and therefore did not align with the evaluation parameter to review two outputs for each project within a grant. After these adjustments, the total pool consisted of 111 grants across nine NIDDR program mechanisms. It is possible that the older grants included in the evaluation had an advantage over the grants ending in 2009 because of the additional time for their outputs to have had an impact.

From this pool of 111 grants, 30 grants (27%) were randomly selected for review in the following way. To balance the desire for the sample of grants to represent the nine program mechanisms included in the pool, the sampling was stratified at the program mechanism level as a proportion of all grants in the sampling pool. For example, there were 36 Field Initiated Project (FIP) grants in the sampling pool (see Table 1), which was 32 percent of all of the grants in the sampling pool (n = 111). Therefore, 32 percent of the 30 grants in the sample should be FIPs (n = 10). The 36 FIPS in the sampling pool were numbered 1 through 36 and then 10 FIP grants were randomly selected, using a website that generated random numbers.

Table 2 in the next section shows the number of grants included in the sample by program mechanism. Proportionally, the number of grants sampled in each program mechanism did not reflect the actual proportions of all grants in the larger NIDRR data set (N = 248), but the sampling method did allow for the largest number of grants in the sample to be FIP grants, which was the largest program mechanism in the NIDRR data set.

TABLE 1 NIDRR Grants Ending Between 2006 and 2010, by Program Mechanisms

Program/Funding Mechanism	Number of Grants	Percent of all Grants	Mean Duration of Grant (years)	Total Grant Funding by Program Mechanism (in $)	Percent of Total NIDRR Grant Funding (for Grants Ending 2006–2010)	Number of Grants in Program Mechanism, by Year Ending, with Grants Included in Sampling Pool Highlighted				
						2006	2007	2008	2009	2010
Model Systems Grants										
Burn Injury (MS-Burn)	5	2.02	6.1	7,271,563	2.34	0	0	5	0	0
Traumatic Brain Injury (MS-TBI)	16	6.45	5.6	29,132,862	9.38	0	7	8	1	0
Spinal Cord Injury (MS-SCI)	17	6.85	6.5	33,977,321	10.94	8	9	0	0	0
Center Grants										
Rehabilitation Engineering and Research Centers (RERC)	12	4.84	5.7	55,816,980	17.98	0	0	0	8	4
Rehabilitation Research and Training Centers (RRTC)	21	8.47	5.9	82,920,345	26.71	0	0	0	10	11
Research and Development Grants										
Disability and Rehabilitation Research Projects (DRRP)	18	7.26	5.0	30,627,386	9.87	0	0	0	14	4
Field Initiated Projects (FIP)	74	29.84	3.8	35,881,454	11.56	0	0	0	36	38
Small Business Innovation Research, Phase I (SBIR)	31	12.50	0.6	2,323,305	0.75	0	0	0	16	15
Small Business Innovation Research, Phase II (SBIR)	16	6.45	2.5	7,990,171	2.57	0	0	0	8	8
Translation Grants										
DRRP-Disability Business Technical Assistance Centers (DBTAC)	1	0.40	1.8	1,742,400	0.56	0	0	1	0	0
DRRP-Knowledge Translation (DRRP-KT)	3	1.21	5.0	8,179,933	2.64	0	0	0	0	3
Training Grants										
Advanced Rehabilitation Research Training (ARRT)	11	4.44	5.8	8,229,338	2.65	0	0	0	1	10
Switzer Fellowships	20	8.06	1.3	1,220,000	0.39	0	0	0	12	8
DRRP-Section 21	3	1.21	6.1	5,141,955	1.66	0	0	1	1	1
Total	**248**	**100.00**		**$310,455,013**	**100.00**	**8**	**16**	**15**	**107**	**102**
Number of Grants in Sampling Pool by End Year						**0**	**9**	**13**	**89**	**0**
Total Number Grants in Pool						**111**				

SOURCE: Data summarized from National Institute on Disability and Rehabilitation Research (September 2009). *Annual Performance Report Data Set of Grants Ending in 2006 to 2010*. Washington, DC: National Institute on Disability and Rehabilitation Research.

After the proposed evaluation methods received approval from the institutional review board of the National Academies, the sample of 30 grants was drawn, and invitations to participate were sent to the principal investigators (PI) of the 30 grants. The PIs were fully informed about the methods to be used in the evaluation and what would be required of them. Of the original 30 grantees invited, 3 (1 DRRP and 2 FIPs) declined because they did not have time to fulfill the evaluation requirements (n = 2) or changed institutions (n = 1). Three other grants were then randomly selected from the remaining pool for the appropriate program mechanisms to bring the final sample to 30 grants (i.e., 1 DRRP and 2 FIPs were drawn). In replacing three of the originally sampled grants, we acknowledge that bias from self-selection could have crept into the evaluation findings and that the final sample of 30 grants that participated in the evaluation may not be fully representative of the larger population of grants.

Compiling Outputs to Be Reviewed and Number of Outputs Reviewed

As noted, the PIs of the grants included in the sample were provided with written instructions about how to submit their outputs for the review and provide supplemental information about the outputs. Committee staff worked with the grantees to clarify the instructions and to encourage them to submit their output packages. Because some grants had ended several years before our review (2007 and 2008 for the Model Systems grants), some grantees had difficulty in submitting materials because the PIs had changed departments or institutions or had other competing priority activities during the time period of our review. Staff accommodated these PIs by providing additional time for submitting their materials and, in five cases, by assisting them in completing the questionnaires through telephone interviews. Two grantees did not provide the supplemental questionnaires.

As described above, grantees were sent questionnaires on which they were asked to list each project under their grant and nominate two outputs per project to be reviewed by the committee. They were asked to identify the top two outputs per project that reflected their grant's best achievements. In order to permit assessment of outputs beyond journal publications, grantees were asked to offer at least one nonjournal publication per project, if such outputs were available. The number of projects for each grant varied by size, from 1 for small field-initiated grants to 10 on larger center grants. Therefore, the number of outputs nominated for review per grant ranged from 2 to 20; the average number of outputs per grant was 5. A total of 156 outputs were submitted for review across the 30 grants selected. Eight outputs were considered highly related to other outputs, and they were reviewed together. This occurred when one output was a derivative or different expression of another output, and when the PI responses to criteria questions were basically the same. Therefore, the number of outputs for analysis was 148. Table 2 presents the number of grants included in the sample by program mechanism and the types of outputs that were reviewed.

To put the outputs reviewed into the larger context of the outputs produced by grantees in the sampling pool of 111 grants, Table 2 also shows that the proportion of publications and other outputs (tools, technology, and information products) that were reviewed by the committee were relatively close to the proportions of the various output types produced by grantees in the larger sampling pool. The proportion of publications reviewed was somewhat lower at 70 percent (compared with 76 percent in the sampling pool), and the proportion of information products reviewed was somewhat higher at 18 percent (compared with 11% in the sampling pool). The mean number of outputs per grant in the sample is much lower (mean = 5) than in the sampling pool (mean = 13) because the sampled grants only submitted their top two outputs per project (as described above).

TABLE 2 Number of Grants and Distribution of Outputs Reviewed, by Program Mechanism

NIDRR Grant Category and Program Funding Mechanisms	Grants	Publica-tions	Tools	Tech-nology	Infor-mation	Total
Model Systems Grants						
Burn Injury (MS-Burn)	2	12	2	0	4	18 (12%)
Traumatic Brain Injury (MS-TBI)	2	12	0	0	2	14 (10%)
Spinal Cord injury (MS-SCI)	2	11	0	0	0	11 (7%)
Center Grants						
Rehabilitation Research and Training Center (RRTC)	3	16	0	0	12	28 (19%)
Rehabilitation Engineering and Research Centers (RERC)	2	16	2	5	3	26 (18%)
Research and Development Grants						
Disability and Rehabilitation Research Projects (DRRP)	4	13	4	0	5	22 (15%)
Field Initiated Projects (FIP)	10	17	1	3	1	22 (15%)
Small Business Innovation Research, Phase II (SBIR)	2	1	0	1	0	2 (1%)
Training Grants						
Switzer Fellowship	3	5	0	0	0	5 (3%)
Total and Proportion of Output Types in Sample	**30**	**103 (70%)**	**9 (6%)**	**9 (6%)**	**27 (18%)**	**148**
Total and Proportion of Output Types in Sampling Pool	**111**	**1,060 (76%)**	**101 (7%)**	**84 (6%)**	**148 (11%)**	**1,393**

SOURCE: Data summarized from Questionnaires submitted to committee by NIDRR Grantees that participated in the evaluation (Rows 3 to 16); and National Institute on Disability and Rehabilitation Research (September 2009). *Annual Performance Report Data Set of Grants Ending in 2006 to 2010*. Washington, DC: National Institute on Disability and Rehabilitation Research (Row 17).

The Review Process

The committee members, whose expertise covers social sciences, rehabilitation medicine, engineering, evaluation, and knowledge translation, were divided into three subgroups of five members each. The subgroups were organized to ensure that outputs would be reviewed by a group of individuals with the collective expertise necessary to judge their quality. The subgroups met in October 2010, December 2010, and February 2011. Because of the relatively short period of time in which to conduct the reviews, grants were scheduled for review according to size, with the smaller grants being invited first (e.g., FIPS, Switzers, SBIRs), and the larger grants (DRRPs, models systems, center grants) being invited to participate in the later rounds. The rationale for the scheduling was that the smaller grants had fewer outputs and would need less preparation time for the review than the larger grants, which had many projects and more outputs to prepare for the review. As a result of this approach, the content of the grants being reviewed in each round tended to be mixed and so required a corresponding mix of expertise in each subgroup. However, efforts were made to match the expertise of the reviewers in each subgroup with the outputs they would be reviewing (e.g., technology output was assigned to a subgroup with engineering expertise). For a detailed description of the review procedures, see Box 2.

BOX 2
Committee Review Procedures

Each of the 30 grants was assigned to one of the three committee subgroups, so that all outputs from a grant were reviewed by the same subgroup. To ensure consistency in approach across subgroups, the committee chair attended all subgroup meetings.

Based on direct review of the output itself and supplemental information about the output provided in the APRs, final reports, and questionnaire responses from grantees, each subgroup member independently rated every output assigned to that subgroup, assigning a quality criteria score for each of the four quality criteria (technical quality, advancement of knowledge or the field, likely or actual impact, and dissemination), as well as an overall score for the output and a rationale for the overall score. Scores were assigned using a 7-point scale ranging from 1 to 7 and anchored at 3 points: 1 = poor quality, 4 = good quality, and 7 = excellent quality.

For each output, one subgroup member was assigned as the primary reviewer; the remaining four subgroup members were secondary reviewers.

The subgroups used the following process for arriving at consensus scores:

- The primary reviewer opened discussion of each output by presenting a brief summary of the output and his or her rationale for rating each relevant criterion plus the overall score.
- The secondary reviewers then presented their ratings for each output and a brief rationale.
- The subgroup then developed consensus group ratings for each output through discussion facilitated by the subgroup chair.

Following the discussion of all outputs from an individual grant, the subgroup considered the full spectrum of the reviewed material, along with the grant's overall purpose and objectives (using the grant's APR), and assigned an overall performance rating for the grant using the same 7-point scale.

The committee's expert review involved a qualitative consideration and assessment of the multiple quality dimensions of the outputs — a process that has been recommended as a valid method for evaluating the relevance and quality of federal research programs (Committee on Science, Engineering, and Public Policy, 1999). The 7-point rating scale was used in order to more precisely describe the results of the output assessment in terms of varying levels of quality. During the reviews, the committee members frequently discussed how they were applying the criteria and interpreting the anchors of the rating scale so they could calibrate their ratings. In addition, brief narrative statements were written that summarized the rationale for the subgroups' ratings of each output. These statements were reviewed after the ratings were completed to identify attributes that particularly characterized the varying levels of quality and were helpful in further exemplifying the dimensions of the criteria.

Although the final scores used to report results of the output assessment were based on the consensus scores, the committee conducted an interrater reliability analysis of their initial independent ratings (i.e., raw scores before their discussion) to determine the degree to which individual committee members were using and interpreting the scale in the same way. The interrater reliability analysis was conducted, using methods suggested by MacLennan (1993), for more than two raters with ordinal data. This method calculates an intraclass correlation coefficient (ICC) that represents an average correlation among raters. The interrater reliability analyses were run on 15 grants that had at least 3 outputs reviewed by the subgroups. The ratings compared were the individual committee members' raw scores (before discussion) on the technical quality criterion only. The ICCs ranged between .64 and .98 and were statistically significant at $p \leq .05$. According to Yaffee (1998), the minimum acceptable ICC is .75 to .80. Of the 15 grants, 13 had ICCs great than .75. The ICC results suggest that individual members were

using and interpreting the 7-point scale in a similar manner prior to the full subgroup's discussions of the output ratings and their subsequent determination of consensus scores.

ASSESSMENT OF THE COMMITTEE'S REVIEW METHODS

The committee developed and implemented an evaluation process for assessing the outputs of NIDRR's grantees and was able to identify varying levels of quality as well as some of the output characteristics associated with these varying quality levels. Considerable time was spent selecting and refining the criteria used to assess quality. Although there was some variation in the independent scoring among subgroup members, it was rarely extreme, particularly after the group discussions. And although the specific content area expertise to assess every output could not be ensured for the diversity and breadth of the outputs reviewed, the committee concludes that, collectively, the subgroups were able to adequately assess all the outputs.

The committee endeavored to assess its evaluation methods throughout the study process. Members engaged in continuous reflection and recording of strengths and weaknesses during the rating process conducted in subgroup meetings. To facilitate this effort, the committee chair participated in all subgroup meetings to ensure members understood how each subgroup was applying the rating methods. In addition, conference calls with the full committee were held after each set of subgroup meetings to discuss the evaluation process and refine the methods. Lastly, during its final meeting, the committee devoted a half-day session to discussion of the strengths and weaknesses of the process and developing conclusions and recommendations for future evaluations. This discussion was based on the continuous reflections of committee members, along with findings from an informal, anonymous poll of committee members about the review process.

In the poll, each committee member was asked to rate his or her level of confidence in 16 aspects of the review process and 8 topics related to its replication. For each of these aspects, members assigned a confidence rating on a 5-point scale in which 1 indicated "no confidence at all" and 5 indicated "extreme confidence." The poll was intended to provide an indicator of each committee member's assessment of the output rating process. Poll results confirmed that individual members were generally confident in the review process and the potential replication of the process, with confidence ratings above the midpoint for all but one of the review process aspects and all but one of the replication topics.

Aspects of the review process in which the committee had the greatest confidence (with scores above 4 on the rating scale) were:

- the technical quality score,
- the face validity of the consensus scores that were produced for outputs,
- the ability of the committee to evaluate outputs without having consumers on subgroups, and
- the appropriateness of a 7-point quality rating scale.

These results were consistent with the committee's overall impressions on the strengths and weakness of the evaluation process over the course of its work.

With regard to the poll item on the ability of the committee to evaluate outputs without having consumers on subgroups, the committee notes that its confidence rating on this item is not meant to suggest that the input of individuals with disabilities is not a necessary part of the process. The committee included two subject-matter experts who are also individuals with disabilities, and the point above relates to committee members' view that the subgroups, while

lacking consumers without relevant scientific expertise, did assign appropriate scores to the outputs.

The poll also confirmed committee impressions regarding the challenge of rating outputs other than peer-reviewed journals; this was the one aspect of the review process receiving an average confidence rating below the midpoint of the scale.

The results of the poll related to replication of the review process largely mirrored the results related to the review process itself. Committee members expressed the greatest confidence in the ability to match appropriate reviewer expertise with outputs to review and the ability to appropriately secure knowledgeable reviewers. The only issue that received an average confidence rating below the midpoint was the ability to assess the overall quality of grants by reviewing selected outputs.

Overall, members' reflections on the summative evaluation process suggest that it worked well and achieved what it was designed to do. However, the committee encountered several challenges and limitations during the course of our work that limit the generalizability of the findings from this evaluation and restrict what can be said about the totality of outputs generated by all NIDRR grantees. In the next section, within the context of recommendations for future evaluations, we discuss these limitations and issues.

RECOMMENDATIONS FOR FUTURE EVALUATIONS

The committee offers conclusions, recommendations, and suggestions on defining evaluation objectives, strengthening the output assessment, and using NIDRR's APR system to capture data for future evaluations. The goal of our recommendations and suggestions is to improve future evaluation efforts and to ensure that evaluation results optimally inform NIDRR's efforts to maximize the impact of its research grants.

Defining Future Evaluation Objectives

The primary focus of the committee's summative evaluation was to assess the quality of research and development outputs produced by grantees. This evaluation did not allow for an in-depth examination or comparison of the larger context of the funding programs, grants, or projects in which the outputs were produced. Although capacity building is a major thrust of NIDRR's center and training grants, assessment of training outputs, such as the number of trainees moving into research positions, was not part of our charge.

NIDRR's grant mechanisms or programs vary substantially in both size and duration (see Table 1, above), with grant amounts varying from less than $50,000 (Field Initiated Projects) to more than $4 million (Center Grants), and grant durations varying from less than 1 year to more than 5 years. Programs also differ in their objectives, so the expectations of the grantees under different programs vary widely. For example, a Switzer training grant is designed to increase the number of qualified researchers active in the field of disability and rehabilitation research. In contrast, Center Grants and Model Systems have multiple objectives that include research, technical assistance, training, and dissemination. Model Systems have the added expectation of contributing patient-level data to a pooled set of data on the targeted condition (i.e., Burn, TBI, SCI).

The number of grants to be reviewed was set at 30 by the committee's charge; this represented about one-quarter of the pool of 111 grants from which the sample was drawn, with the requirement that the sample reflect grants across NIDRR's program mechanisms. Even though five program mechanisms were not included in the sampling pool, the number of grants

reviewed for any of the remaining nine program mechanisms was very small. (The largest number of grants reviewed for any single program mechanism was 10—for FIPs). Since the number of grants reviewed for any given program was small, the committee did not attempt to make comparisons of the type or quality of outputs by program mechanism.

The committee was directed by NIDRR to review two outputs for each of the projects identified by a given grantee. Therefore, a grantee with a single project had two outputs reviewed, a grantee with three projects had six outputs reviewed, and so on. Although larger grants with more projects also had more outputs reviewed, the current design considers neither grant size nor duration. The design also did not take into consideration the relative importance of a given project within a grant.

The committee was also asked to produce an overall grant rating based on the outputs reviewed and the information available about the grants from the APRs. Results at the grant level are subject to more limitations than those regarding outputs due to the general lack of information about how the outputs did or did not interrelate; whether, and if so, how grant objectives were accomplished; and the relative priority placed on the various outputs. In addition, for larger, more complex grants, such as Center Grants, a number of grant expectations, such as capacity building, dissemination, outreach, technical assistance, and training, are unlikely to be adequately reflected in the approach used, which focused exclusively on specific outputs. The relationship of outputs to grants is more complex than this approach allowed.

Recommendation 1: NIDRR should determine whether assessment of the quality of outputs should be the sole evaluation objective.

Considering other evaluation objectives might offer NIDRR further opportunities to continuously assess and improve its performance and achieve its mission. Alternative designs would be needed to evaluate the quality of grants or to allow comparison across program mechanisms. For example, if one goal of an evaluation is to assess the larger outcomes of grants (i.e., the overall impact of a grant's full set of activities), in addition to the methods used in the current output assessment, the evaluation would need to include interviewing grantees about their original grant objectives, to learn about how the grant was implemented and any changes that may have occurred in the projected pathway, how various projects were tied into the overall grant objectives, and how the outputs demonstrated the achievement of the grant and project objectives. This approach would also involve conducting bibliometric or other analyses of all publications and examining documentation of the grant's activities and its self-assessments, including cumulative APRs over time. Focusing at the grant level would provide evidence of movement along the research and development pathway (e.g., from theory to measures, from prototype testing to market), as well as allowing for assessment of other aspects of the grant, such as training and technical assistance and the possible synergies of multiple projects within one grant.

If the goal of an evaluation is to assess and compare the impact of program mechanisms, different methods may be needed, depending on the expectations for each program mechanism. They would need to include not only those mentioned above, but also stakeholder surveys to learn about the specific ways that individual grants affect their intended audiences. And in order to allow for generalization and comparison across program funding mechanisms, larger grant sample sizes would be needed. An alternative would be to increase the grant sample size in a narrower area by focusing assessments on

grants for specific research areas across different program mechanisms or on grants with shared objectives (e.g., product development, knowledge translation, capacity building).

NIDRR's questions will necessarily drive future evaluations, but other levels of analysis that NIDRR might focus on could include the portfolio level (e.g., Model System grants, research and development, or training grants), which NIDRR has done in the past; the program priority level (i.e., grants funded under certain NIDRR funding priorities) to answer questions regarding the quality and impact of NIDRR's priority setting; and institute-level questions to evaluate the net impact of NIDRR grants or to test assumptions embedded in NIDRR's logic model. For example, NIDRR's intermediate outcome arena targets adoption and use of new knowledge leading to changes/improvements in policy, practice, behavior, and system capacity (see *Federal Register*, February 15, 2006, pp. 8,173–8,175).

The number of outputs reviewed should depend on the unit of analysis. At the grant level, it might be advisable to assess all outputs to examine their development, how they relate to one another, and their impacts. A case study methodology could be used for subsets of outputs that are related. If NIDRR aims its evaluation at the program funding mechanism or portfolio level, sampling grants and assessing all outputs would be the preferred method. For output-level evaluation, having grantees self-nominate their best outputs, as was done in the present evaluation, is a good approach.

Although assessing grantee outputs is of great value, it is the committee's view that the most meaningful results would come from assessing outputs in the context of a more comprehensive grant-level evaluation. More time and resources would be required to trace a grant's progress over time in accomplishing its objectives, to understand its evolutionary development that might have altered original objectives, and to examine the specific projects that produced the various outputs. However, more closely examining the inputs and processes of grant implementation that produced the outputs would yield broader implications for the value of grants, their impact, and future directions for NIDRR.

Strengthening Future Output Assessments

The committee was able to create a reasonably reliable system for evaluating the outputs of NIDRR grantees based on criteria used in assessing federal research programs both in the United States and other countries. With refinements, it could be applied to evaluate future outputs even more effectively. In implementing the output-level assessment, particular challenges and issues arose in relation to the diversity of outputs, the timing of evaluations, sources of information, and reviewer expertise.

Diversity of Outputs

The quality rating system used in the summative evaluation worked very well for publications in particular, which comprised 70 percent of the outputs reviewed. Using the four criteria developed by the committee, the reviewers were able to identify varying levels of quality and the characteristics associated with each of them. However, each of the quality criteria was not so easily applied for diverse outputs such as websites, conferences, and interventions. These outputs require more individualized criteria for assessing specialized technical elements and sometimes more in-depth evaluation methods. Applying one set of criteria, even though broad and flexible, could not guarantee sufficient and appropriate applicability to every type of output.

Timing of Evaluations

The timing of an assessment of outputs depends on the goal of the assessment. Assessing technical quality can be done immediately, but assessing impact of outputs requires time between the release of an output and its eventual impact. Evaluation of outputs during the final year of an award may not allow sufficient time for them to have full impact. For example, some publications will be forthcoming, and others will not have had sufficient time to have an impact. The tradeoff of waiting a year or more after the end of a grant is the likelihood that staff involved with the original grant may not be available, recollection of grant activities may be compromised, and engagement or interest in demonstrating results may be reduced. However, publications can be tracked regardless of access to the grantee. Outputs other than publications, such as technology products, could be assessed in an interim evaluation.

Sources of Information

Committee members were provided with structured briefing books containing the outputs to be reviewed and supplemental information that members could draw on if additional information was needed to assign quality scores. The supplemental information included information submitted through the grantees' APRs and final reports and information provided in a supplemental questionnaire developed by the committee (see Attachment C). The primary source of information used by committee members in assigning scores was direct review of the output itself. The supplemental information played a small role in assessing publications; for outputs such as newsletters and websites, this information could provide needed context and additional evidence helpful in determining quality scores. However, it is important to note that the supplemental information involved grantees' self-reports, which may be susceptible to social desirability bias. Therefore, committee members were cautious in the degree to which this information could serve as the basis for assigning higher output scores. Moreover, the APR was designed for grant-monitoring and performance reporting rather than as a source of information for a program evaluation.

As a supplemental source, the information supplied on the APRs and the questionnaire was not always sufficient to inform the quality ratings. As examples, the technical quality of a measurement instrument was difficult to assess if there was insufficient information about its conceptual base or its development and testing. For conferences, workshops, and websites, it would have been preferable for the grantee to identify the intended audience so that the committee might better assess whether the described dissemination activities were successful in reaching it.

For the output categories of tools, technology, and informational products, grantees sometimes provided a publication that did not necessarily describe the output. In addition, some outputs were difficult to assess when there was no corroborating evidence provided to support grantees' claims about technical quality, advancement of the field, impact, or dissemination efforts.

The committee did not use standardized reporting guidelines, such as CONSORT (Schultz et al., 2010) or PRISMA (Mohrer et al., 2009), which journals use in their peer review processes for selecting manuscripts for publication. The committee members generally assumed that publications that were peer-reviewed warranted a minimum score of 4 for technical quality, which could be changed after the committee's discussion. In some cases, the final committee scores for technical quality for peer-reviewed publications were above 4; in other cases, the final

scores were below 4. If reporting guidelines had been used in the review of research publications, it is possible that the ratings would have changed.

Reviewer Expertise

The committee was directed to assess the quality of four types of specified outputs. Although the most common output type was publications, NIDRR grants produce a range of other complex, varied outputs, including tools and measures, technology devices and standards, and informational products. These outputs vary widely in their complexity and the investment needed to produce them. For example, a newsletter is a more modest output than a new technology or device. To assess the quality of outputs, the committee members used criteria that were based on the cumulative literature reviewed and their own research expertise in diverse areas of rehabilitation and disability research, medicine, and engineering, as well as expertise in evaluation, economics, knowledge translation, and policy. However, the combined expertise of the panel did not include every possible content area in the broad field of disability and rehabilitation research.

Recommendation 2: In any future evaluations of output quality, NIDRR should refine the process developed by the committee to strengthen the design related to the diversity of outputs, the timing of evaluations, sources of information, and reviewer expertise.

Corresponding to the points above these refinements include the following.

Diversity of Outputs The dimensions of the quality criteria should be tailored and appropriately operationalized for different types of outputs, such as devices, tools, and information products (including newsletters, conferences, and websites) and should be field tested with grants under multiple program mechanisms and refined as needed.

For example, the technical quality criterion includes the dimension of accessibility and usability. The questionnaire asked grantees to provide evidence of these traits. However, the dimensions should be better operationalized for different types of outputs. For "tools," such as measurement instruments, the evidence to be provided should pertain to pilot testing and psychometrics. For informational products, such as websites, the evidence should include results of user testing, assessment of usability features, compliance with Section 508 standards (regulations from the 1998 amendment to the Rehabilitation Act of 1973 requiring the accessibility of federal agencies' electronic and information technology to people with disabilities), etc. For technology devices, the evidence should document the results of research and development tests related to human factors, ergonomics, universal design, product reliability and safety, etc.

The quality criterion related to dissemination provides other clear examples of the need for further specification and operationalization of the dimensions. For example, the dissemination of technology devices should be assessed by examining the progress toward commercialization, grantees' partnerships with relevant organizations, including consumers and manufacturers, and the delivery of information through multiple media types and sources tailored to intended audiences for optimal reach and accessibility.

Timing of Evaluations The committee suggests that the timing of an output assessment should vary by output type. Publications would best be assessed at least 2 years after a

grant ends. However, plans for publications and other dissemination, as well as the audience for scientific papers, could be included as an item in the final report. As discussed above, other outputs developed during the course of a grant should be evaluated on an interim basis to assess the development and evolution of products. Outputs that have the potential to affect practice or policy may require longer periods of time to pass before impact materializes and can be measured, so they would also best be evaluated on an interim basis.

Sources of Information A more proactive technical assistance approach is needed to ensure that grantees provide the data necessary to assess the specific dimensions of each quality criteria**.** As stated above, the information supplied on the APRs and the questionnaire was not always sufficient to inform the quality ratings. (See also the discussion of information requested in the grantee questionnaire, above, and the discussion of APRs, below.)

Reviewer Expertise The committee suggests that future output evaluations should consider including an accessible pool of experts in different technical areas who can be called on to review selected grants and outputs. In addition, it is essential that future review panels include scientists with disabilities. Consumers, who are not scientists, could also play a vital role as review panel members who can address the impact and dissemination criteria.

Using Annual Performance Reports for Evaluation

NIDRR's APR system has numerous strengths, but the committee identified some points that NIDRR should consider in building greater potential for use of these data in evaluations. The APR system (Research Triangle Institute, 2009) includes the grant abstract, funding information, descriptions of the research and development projects, and outcome domains targeted by projects, as well as a range of variables for reporting on the four different types of grantee outputs; see Table 3. The system is tailored to different program mechanisms as needed. All of the descriptive information listed above, plus the output-specific variables listed in Table 3, were used in the committee's work. The data were provided to the committee as electronic data bases and in the form of individual grant reports.

The APR data provided to the committee by NIDRR at the outset of our work was used to profile the grants for sampling and in listing all of the grantees' projects and outputs. They facilitated asking the grantees to nominate outputs for our review, since we were able to generate comprehensive lists of all reported projects and outputs to make the task of output selection less burdensome for the grantees. If grantees had more recent outputs that they wished to nominate as their top two for the committee's review, they had the option to do so.

TABLE 3 Data Elements Related to Outputs That Are Covered in an APR

Variables in APR[a]	Publications	Tools	Technology	Information
Type of output	X	X	X	X
Name and full citation	X	X	X	X
Brief description of purpose		X	X	X
Brief description of how output was validated or tested		X	X	X
Whether publication was peer reviewed or not	X			
Whether the research and related activity reported in the article took place during current, immediate past, or previous	X			

Variables in APR[a]	Publications	Tools	Technology	Information
(nonconsecutive) funding cycle				
Whether publication was sent to NARIC for inclusion in REHABDATA	X			
Whether publication was produced as a direct result of receiving funding for this grant?	X			
"Most important"[b] outputs that contributed the most to achieving the outcome-oriented goals for the award	X	X	X	X
Outcome-oriented goal that corresponds to most important outputs (advances knowledge; increases capacity for research, training, or knowledge translation; or facilitates change in policy, practice, or system capacity)	X	X	X	X
NIDRR outcome arena that corresponds to most important outputs (health and function, employment, participation and community living, cross-cutting)	X	X	X	X
Whether output is described in a publication output and indicate which one		X	X	X
Key findings or lessons learned	X			
How output is contributing to the outcome-oriented goal by solving a problem, closing an identified gap, or benefiting the target population	X	X	X	X

[a]SOURCE: Using NIDRR APR report format for Rehabilitation Research and Training Centers as an example
[b]Defined for grantees by NIDRR as "those that contributed most to achieving the outcome-oriented goals for the award by advancing knowledge, increasing capacity for research, training or knowledge translation; or facilitating changes in policy, practice, or system capacity."

NIDRR also provided grantees' narrative APRs from the last year of the grants, as well as their final reports. These narratives were very useful to the committee for compiling descriptions of the grants. However, the quality of the information contained in the narrative annual reports varied.[3] For example, grant abstracts were not uniform in the information they contained. Some stated their grant objectives; others omitted them and focused on summarizing their main grant activities. The APRs of the grants reviewed were inconsistent in providing useful information for understanding how the outputs being reviewed fit in the context of the overall grant or projects. The final reports in most cases did not provide a cumulative overview of the life cycle of the grants and outputs, which would have been helpful. The APR does collect information on changes in the course of grants, but it was not always easy to understand this information from just viewing the last year's APR or the final report.

NIDRR also provided the committee with special text reports that contained some of the narrative information in the APRs about outputs other than publications. These reports included such information as the purpose of the output, NIDRR outcome domains targeted by the output, how the output was validated, and how the output contributes to achievement of the grantee's goals. These reports have the potential to supply contextual information for evaluations. However, the quality of the information in them varied across the text reports describing the tool, technology, and information outputs that the committee reviewed. Only half of the text reports contained substantive descriptive information.

[3]The APR is a large information technology system that is used for monitoring and tracking grantee progress and for reporting on NIDRR's performance measures under the Government Performance and Results Act (GPRA). The system was not designed to serve as the basis for grantee evaluations. A systematic evaluation of the APR was not part of our charge. Though the quality and level of detail included in the APRs varied, these narratives were useful in providing descriptive grant information.

Not all of the specific outputs reviewed by the committee were reported in the APRs. Some may have been reported in earlier reporting periods or had been produced after the NIDRR grant ended.

Recommendation 3: NIDRR should consider revising its APR to better capture information needed to routinely evaluate the quality and impacts of outputs, grants, and funding mechanisms. They might consider such efforts as consolidating existing data elements or adding new elements to capture the quality criteria and dimensions used in the committee's summative evaluation.

From a recent interview with senior executives at NIDRR, the committee learned that NIDRR takes pride in having stabilized its APR system in recent years after prior periods of changing and improving it to make the data more usable for grantees, for grant monitoring by project officers, and for agency performance reporting. We were informed that NIDRR is currently in the process of adding a new "accomplishments" module to the APR that will focus on the external use and adoption of NIDRR-funded outputs. In this new module, NIDRR will consolidate some data elements that are already being collected and add new ones. For up to five outputs that have been used or adopted by persons or groups external to the grant during the reporting period, grantees will be asked to provide information for each output on who adopted the outputs (in 16 categories, such as researchers, practitioners, service providers); how the output is being used or adopted by the target audience; the source of the evidence; and if and how the output may be contributing to changes in policy, practice, system capacity, or other impact areas. These efforts that are under way to change the APR will address the quality criteria used in the committee's evaluation for assessing the advancement of knowledge or practice and the likely or demonstrated impact of outputs.

For the technical quality criterion, the current APR system collects data on whether articles were published in peer-reviewed journals. For the technical quality of outputs other than publications, we provide examples in the discussion of Recommendation 2 (above) of ways to operationalize dimensions of accessibility and usability, such as providing evidence of testing the psychometrics of measurement instruments, assessing the usability features of informational products, and documenting the results of research and development tests of technology products that relate to human factors, ergonomics, universal design, product reliability, and safety. The APR system currently asks for information on how outputs were validated, but data elements that relate to such testing might be further specified in the APR system.

The APR system might also be modified to capture evidence on the quality criterion of dissemination of outputs through such data elements as target audiences for dissemination activities, media types, number of outputs disseminated, and reach of dissemination, such as number of hits on websites.

Recommendation 4: NIDRR should investigate ways to work with grantees to ensure the completeness and consistency of information provided in the APRs.

The committee fully appreciates the necessity of minimizing the data collection burden on grantees and acknowledges the challenges and feasibility issues related to

modifying the APR system while at the same time providing continuity in the system. The committee suggests, however, that embedding evaluation data collection processes into existing processes will lead to greater efficiencies and reduce grantee burden while enhancing NIDRR's ability to evaluate quality and impact. The committee acknowledges that the refinements suggested would have to be undertaken in the context of a larger assessment of the APR system as part of NIDRR's ongoing initiatives to improve the system.

In sum, the committee was able to create a reasonably valid and reliable system for evaluating the outputs of NIDRR grantees. If future evaluations of output quality are conducted, the process developed by the committee should be implemented with refinements to strengthen the design and process. Although assessing grantee outputs is of great value, the committee thinks that even greater value would come from assessing outputs in the context of a more comprehensive grant-level evaluation, which could yield broader implications for the value of grants, their impact, and future directions for NIDRR.

The committee has appreciated the opportunity to work on this important endeavor, and we look forward to delivering our final report to you later this year.

Sincerely yours,

David H. Wegman, *Chair*
Committee on the External Evaluation of NIDRR and Its Grantees

REFERENCES

Bernstein, A., Hick, V., Borbey, P., Campbell, T., McAuley, L., and Graham, I.D. (2007). A framework to measure the impacts of investments in health research. In Organisation for Economic Co-operation and Development (Eds.), *Science, technology and innovation indicators in a changing world: Responding to policy needs* (pp. 231-251). Ottawa: Organisation for Economic Co-operation and Development.

Canadian Academy of Health Sciences. (2009). *Making an impact: A preferred framework and indicators to measure returns on investment in health research.* Appendices. Panel on Return on Investment in Health Research. Ottawa: Author.

Chien, C.F., Chen, C.P., and Chen, C.H. (2009). Designing performance indices and a novel mechanism for evaluating government R&D project. *Journal of Quality, 19*(2), 119-134.

Committee on Science, Engineering, and Publc Policy. (1999). *Evaluating federal research programs: Research and Government Performance and Results Act*. Washington, DC: National Academy Press.

Federal Register. (2006). *Department of Education. National Institute on Disability and Rehabilitation Research—Notice of Final Long-Range Plan for Fiscal Years 2005–2009*, Volume 71, No. 31, Feb. 15.

Ismail, S., Tiessen, J., and Wooding, S. (2010). *Strengthening research portfolio evaluation at the Medical Research Council: Developing a survey for the collection of information about research outputs.* Santa Monica, CA: RAND.

MacLennan, R.N. (1993). Interrater reliability with SPSS for Windows 5.0. *American Statistician, 47*(4), 292-296.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., and The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*. Available: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000097 [June 2011].

National Institute on Disability and Rehabilitation Research. (2008, August). *Meeting report of the NIDRR External Evaluation Advisory Group.* Washington, DC: Author.

National Institute on Disability and Rehabilitation Research (NIDRR). (2009). *Briefing book for The National Academies*. (Resources and documentation provided by NIDRR to The National Academies). Washington, DC: Author.

National Research Council. (2007). *Mining safety and health research at NIOSH.* Committee to Review the NIOSH Mining Safety and Research Program. Report. No. 2, Reviews of research programs of the National Institute for Occupational Safety and Health. Washington, DC: The National Academies Press.

Office of Special Education and Rehabilitative Services. (2007). Office of Special Education and Rehabilitative Services: About NIDRR. Available: http:www.ed.gov/about/offices/list/osers/nidrr/about.html [January 2011].

Research Triangle International. (2009). *National Institute on Disability and Rehabilitation Research. Annual Performance Report: Web-based reporting system instruction manual*. Research Triangle Park, NC: Author.

Schulz K.F., Altman D.G., Moher D., for the CONSORT Group. (2010). *CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials*. Available: http://www.consort-statement.org/index.aspx?o=3444 [June 2011].

Wooding, S., and Starkey, T. (2010). *Piloting the RAND/ARC Impact Scoring System (RAISS) tool in the Canadian context.* Cambridge, UK: RAND Europe.

Wooding, S., Nason, E., Starkey, T., Haney, S., and Grant, J. (2009). *Mapping the impact: Exploring the payback of arthritis research.* Santa Monica, CA: RAND.

Yaffee, R.A. (1998). Enhancement of reliability analysis: Application of intraclass correlations with SPSS/Windows v.8. Available: http://www.nyu.edu/its/statistics/Docs/intracls.html [April 2011].

Attachment A

Committee on the External Evaluation of NIDRR and Its Grantees

DAVID H. WEGMAN (*Chair*), School of Health and Environment, University of Massachusetts, Lowell
THOMAS J. ARMSTRONG, Center for Ergonomics, University of Michigan
BURT S. BARNOW, Trachtenberg School of Public Policy and Public Administration, George Washington University
LEIGHTON CHAN, Rehabilitation Medicine Department, Clinical Center, National Institutes of Health
PETER C. ESSELMAN, Department of Rehabilitation Medicine, University of Washington, Seattle
WALTER FRONTERA, University of Puerto Rico School of Medicine
GLENN T. FUJIURA, Department of Disability and Human Development, University of Illinois at Chicago
BRUCE M. GANS, Kessler Institute for Rehabilitation, West Orange, New Jersey
IAN GRAHAM, Knowledge Translation and Public Outreach, Canadian Institutes of Health Research
LISA I. IEZZONI, Mongan Institute for Health Policy, Massachusetts General Hospital, Boston
ALAN M. JETTE, School of Public Health, Boston University
THUBI KOLOBE, Department of Rehabilitation Sciences, University of Oklahoma Health Sciences Center
PAMELA LOPREST, Urban Institute, Washington, DC
KATHRYN E. NEWCOMER, Trachtenberg School of Public Policy and Public Administration, George Washington University
PATRICIA M. OWENS, Government Accountability Office, Minisink Hills, PA
ROBERT G. RADWIN, Department of Biomedical Engineering, University of Wisconsin

JEANNE C. RIVARD, *Senior Program Officer and Co-Study Director* (from September 2010)
MARY ELLEN O'CONNELL, *Co-Study Director* (from September 2010)
MOLLY STORY, *Study Director* (through September 2010)
GARY FISCHER, *Senior Program Assistant*
JATRYCE JACKSON, *Senior Program Assistant* (from September 2010)
MATTHEW D. MCDONOUGH, *Research Associate*
TINA WINTERS, *Associate Program Officer* (from April 2011)

Attachment B

Acknowledgment of Reviewers

This letter report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the Report Review Committee of the National Research Council. The purposes of the independent review are to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards of objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following for their review of the letter report: Stephen H. Bell, senior fellow and principal scientist, Abt Associates, Inc., Bethesda, MD; Susan M. Daniels, president, Daniels & Associates, LLC, Washington, DC; Jody L. Fitzpatrick, School of Public Affairs, University of Colorado Denver; David B. Gray, Health Sciences, Program in Occupational Therapy, Washington University School of Medicine; and John L. Melvin, Michie professor and chair, Department of Rehabilitation Medicine, Jefferson Medical College of Thomas Jefferson University. Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the letter report was overseen by the William C. Howell, Department of Psychology, Rice University and Arizona State University, and John C. Bailar III, professor emeritus, University of Chicago. Appointed by the National Research Council, they were responsible for making certain that an independent examination of the letter report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the letter report rests entirely with the committee and the institution.

Attachment C

This attachment comprises five parts: C-1, the grantee invitation letter; C-2, an informed consent form; C-3, the grantee questionnaire for the summative evaluation; C-4, the committee review procedures for the summative evaluation; and C-5, the rating sheet for committee members for the quality of outputs.

C-1 GRANTEE INVITATION LETTER

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Division of Behavioral and Social Sciences and Education
Committee on Human-Systems Integration

500 Fifth Street, NW
Washington, DC 20001
Phone: 202 334 2678
Fax: 202 334 2210
Email: cohsi@nas.edu
www.nationalacademies.org

October 6, 2010

Grantee Address

Dear Dr. _____:

Last week we sent you an email to inform you that you are being invited to participate in the External Evaluation of the National Institute on Disability and Rehabilitation Research (NIDRR) and Its Grantees that is being conducted by an expert committee of the National Research Council of the National Academies. This independent evaluation is being sponsored by NIDRR for the purpose of: (1) assessing NIDRR's priority setting and peer review processes and (2) reviewing the quality of grantee outputs for a sample of grants that represent the NIDRR portfolio.

Your NIDRR-funded grant (Grant #_____, Grant Title:_____) was selected to be reviewed as part of the evaluation of grantee outputs. However, your participation is completely voluntary. If you do agree to be part of the evaluation, your participation will involve the following activities.

1. We will ask you to nominate two outputs that were produced under each project that was funded by the grant. These will be outputs that best reflect your grant's achievements. We are using the four NIDRR categories of outputs as defined in NIDRR's Annual Performance Report (APR), which include (a) Publications; (b) Tools, Measures, and Intervention Protocols; (c) Technology Products and Devices; and (d) Informational Products. The committee will assess the quality of the outputs which you identify using criteria of technical quality, the extent to which they advance knowledge, their potential impact, and their dissemination.
2. To conduct the review, we would like to examine the actual outputs and to review any documentation that you may have about the outputs.
3. We will ask you to complete a questionnaire about each output that asks you to briefly summarize evidence of their technical quality, how they advance knowledge, their potential impact, and how they were disseminated.
4. The questionnaire will also ask you to respond to a brief set of questions at the grant level about approaches you used in managing your grant, how the grant may have generated new research and projects, and about your perspectives of key NIDRR processes which may influence grant results.

5. We may also ask you to participate in a follow-up telephone or videoconference interview. If so, we would like to audio-record the discussion between you and the committee.

The attachments that follow this letter include, first, an informed consent form that explains what we will do with the results of the evaluation and what steps we will take to protect the confidentiality of the evaluation results specific to your grant. The second attachment is the Grantee Questionnaire referred to above that provides instructions for:

- identifying outputs to be reviewed,
- completing the supplemental questions for each output, and
- sending us your signed consent forms, your outputs, and your completed questionnaire.

Please note that we are sending this package to you in electronic and hard copy form.

We hope that you will decide to participate. If you have any questions please don't hesitate to contact one of us using the information below.

Sincerely,

Jeanne Rivard, Ph.D., Co-Study Director
The National Academies
National Research Council
500 Fifth Street, NW
Washington, DC 20001
Phone: 202-334-2967
Email: jrivard@nas.edu

Mary Ellen O'Connell, Co-Study Director
The National Academies
National Research Council
500 Fifth Street, NW
Washington, DC 20001
Phone: 202-334-2607
Email: moconnell@nas.edu

C-2 INFORMED CONSENT FORM
(For NIDRR Grantees)

What the study is about: An expert committee of the National Research Council of the National Academies, in Washington, DC is developing an evaluation framework that will be used to: 1) review NIDRR's priority setting and peer review processes; and 2) review the quality of grantee "outputs" for a sample of grants that represent the NIDRR portfolio. ["Peer review" refers to a process in which experts review the merits of a grant application in considering whether it should be funded. "Outputs" are publications, measures, intervention protocols, devices, or information resources that are produced as part of a grant.] The committee will also assess the design and implementation of the evaluation process and make recommendations for additional evaluation cycles that may be performed subsequent to this effort.

What we will ask you to do: We would like to invite you to participate in the evaluation.

Your participation will involve:

1. Having outputs produced under your grant peer reviewed through a quality assessment by an expert panel.
2. To conduct the review, we would like to examine a copy of the actual outputs and to review any documentation that you may have about the outputs.
3. We will ask you to complete a questionnaire about the outputs to assist in assessing their technical quality, the extent to which they advance knowledge, their potential impact, and if applicable, their dissemination.
4. We may ask you to participate in a follow-up telephone or videoconference interview where we would audio-record the discussion between you and the committee.
5. We will also ask you to respond to a brief set of questions at the grant level about your grant management, the generation of new research and projects, and about key NIDRR processes which may influence grant results.

Taking part is voluntary. Your participation is completely voluntary. You can chose not to answer some of the questions, and there will be no consequences.

Benefits and risks:

Benefits: By taking part in the evaluation you will provide information that may help NIDRR improve its research portfolio for the benefit of persons with disabilities.

Risks: Because NIDRR has funded some of your research and development activities, you may feel uncomfortable having your grant's outputs formally rated in the study, or in sharing your opinions and perspectives on NIDRR's key management processes. You might feel that this could be a risk to your future grant funding. We want you to know, however, that we will take every step necessary to protect your confidentiality and minimize this risk.

Your answers will be confidential. In the final evaluation report, we will briefly describe your grant and the outputs that were reviewed, by their titles and grantee institutions. Your name as the Principal Investigator, or other investigators on your team, will not be used. In most cases we expect that you have already placed information about the outputs in the public domain through publications, presentations at conferences, and through NIDRR's National Rehabilitation Information Center (NARIC) website.

However, distinct from these descriptions will be your responses on the Grantee Questionnaire and the committee's quality ratings of your outputs. For analysis and reporting, these narrative and quantitative

data will be de-identified and aggregated across all outputs and all grants. A research identification number will be used to track grants and their specific outputs. Outputs may be analyzed by categories, such as output type (e.g., publications, tools, technology, information products); quality criteria assessed (e.g., technical quality, knowledge advancement, potential impact); or program funding type (e.g., center grant, field initiated grant, training grant, etc.). If your grant or your output represents one of a kind and there is a risk of identifying you because of this, your data will be aggregated with another larger group where identification will not be a risk.

Every effort will be made to protect the confidentiality of the information that you provide. The Study Director will keep a list linking the grant and output research ID numbers with that output's identifying information (institution, grant title). This list, along with the data collected, will be stored securely at the National Research Council, and will be accessible only by the Study personnel. If a telephone or videoconference interview is convened to gather additional follow-up information, the transcription of audio-recorded interviews will be combined in a dataset with the interviews of all of the other respondents, then analyzed for common themes across the interviews. The audiotapes, transcriptions, grantee questionnaires, committee ratings, and other raw data collected will be destroyed at the end of the study when the report is released.

Compensation: There is no compensation for participating in the evaluation.

If you have any questions: The Co-Study Director of the evaluation and contact for questions is Jeanne Rivard, Ph.D. If you have any questions about this consent form or the study, she can be contacted by phone at: 202-334-2697, or by email at: jrivard@nas.edu.

If you have any questions, comments, or concerns about taking part in this study, first talk to Dr. Rivard above. If for any reason you do not want to do this, or you still have concerns after doing so, you may contact the Institutional Review Board (IRB) of the National Academy of Sciences (NAS), which reviewed and approved the study plans and this consent form. You can reach the chair of the IRB by contacting Ronald D. Taylor, Human Protections Administrator, by telephone at 202-334-1659 or you may write to him at the National Academy of Sciences; Room 1026; 500 Fifth Street, NW; Washington, DC 20001.

Statement of Consent: I have read the above information, and have received answers to any questions I asked. I consent to take parting in the study.

Your Signature_____ Date__-__-__

Your Name (printed) _____

In addition to agreeing to participate, if there is a follow-up interview, I also consent to having it tape-recorded.

Your Signature_____ Date__-__-__

C-3 SUMMATIVE EVALUATION: GRANTEE QUESTIONNAIRE

Grant Award Number:
Grant Title:
Grantee:
Program Mechanism:
Grant End Date:

INSTRUCTIONS TO GRANTEES

This questionnaire has been designed to obtain information to assist the Committee in assessing the quality of your grant's outputs. NIDRR has provided to the Committee and the National Research Council (NRC) staff copies of its Annual Performance Report (APR) database and your last APR and your final APR. As you will see in certain places on the questionnaire we have inserted information from your APR to facilitate your completion of the questionnaire (e.g., Table 1 lists your research and development projects, and Table 2 lists your outputs reported in the APR). Where this information is in error, we would appreciate your pointing the errors out to us and correcting it; and/or updating the information as needed.

The questionnaire is divided into the following three parts:

Part A. **Nominating Outputs for Review.** This section asks you to nominate, for the Committee's review, the "top 2" outputs for each of your projects that best reflect your grant's achievements (Table 1 below). The Committee would prefer to review one publication and one other type of output for each project. However if you only have publications, please nominate these as your "top 2".

To make this process easier, the NRC staff has populated a list of the outputs (Table 2 below) that were reported for your grant in the APR. You could select the top 2 outputs from this table. However you are not constrained to select from this list if there are other outputs that you think better reflect your grant's achievements.

For Committee review we are requesting materials and information regarding the actual outputs selected as the top 2 for each project.

- For publications, the material for review would be pdf copies of each article.
- For the other outputs, materials for review would include:
  - Electronic or hard copies of the measures, tools, intervention protocols, manuals; or links to websites, pictures or other graphic representations of tools or devices that have been produced.
  - An abstract or summary of each output, which briefly describes:
    - what the output is,

- its purpose,
- target audience,
- methods, and
- how the output fits into the overall goals and objectives of the project and grant

Part B. Additional Questions about Outputs. For each of the outputs you nominated for review, the Committee has a series of questions related to their technical quality, how they may have advanced knowledge, their potential impact, and their dissemination. We ask that you complete the Part B section for each output. If the answers to certain questions would be the same across different outputs, you can note this and cut and paste responses from earlier output forms to other ones. Please make your responses brief, but as specific and quantitative as possible.

Part C. Grant-level Questions. The questionnaire will also contain a few other items asking about how you managed your grants to produce the highest quality outputs, how your grant's results may have generated new projects, and how key NIDRR processes influence results.

Your complete package of materials will contain:

- **Your signed informed consent form**
- **Copies of your publications and other outputs (e.g., measures, tools, intervention protocols, manuals, links to websites, pictures or other graphic representations of devices that have been produced)**
- **Your completed Grantee Questionnaire**

Please send these materials by ___DATE___ to:

Matt McDonough
The National Academies
National Research Council
500 Fifth Street, NW
WS 1134
Washington, DC 20001

We are enclosing an addressed Fedex form that can be used when mailing your package of materials to us. We estimated a shipment cost that would cover a weight up to 10 lbs. (e.g., for large center grants or devices). If your package weighs more than this, Fedex will charge us the correct amount.

If your package is light and you want to send it electronically, you could email it to Matt at mmcdonough@nas.edu. However, you would need to scan your signed consent form, and send that in a pdf document.

Part A. Nominating Outputs for Review

When referring to "outputs", we are using the four NIDRR categories of outputs as defined in NIDRR's Annual Performance Report, which include: a) Publications; b) Tools, Measures, and Intervention Protocols; c) Technology Products and Devices; and d) Informational Products.

Per the instructions for nominating outputs for review, please record your nominations for your "top 2" outputs for each of your projects in Table 1 below. (Reminder "top 2" refers to those that best reflect your grant's achievements). As you can see the NRC staff has already populated Table 1 with the names of your research and development projects from data in the APR. Table 2, which follows, contains a list of outputs from which you can cut and paste into Table 1 below. Please identify any errors in this information that we have provided from your APR and correct it as needed.

Table 1. Projects and Nominated Outputs

#	**Names of R & D Projects in Grant**	**Names of Top 2 Outputs for each Project** *Outputs to be inserted by grantee from Table 2 below or add others as needed*
Research Projects		
R1		1.
		2.
Development Projects		
D1	Dissemination	1.
		2.

INSERT GRANTEES' PROJECTS TABLE HERE

The table below lists all of the publications and other outputs that were listed in the APR data provided by NIDRR. Please use this table below in selecting your top 2 outputs for each project. (You can cut and paste from Table 2 into Table 1.) However you are not constrained to select only from this list if there are other outputs that you think better reflect your grant's achievements.

Table 2. List of Outputs from APR

Type of Output	**Title of Output**
publications	(title)
tool	(title)

Part B. Additional Questions about Outputs

<u>Please use one copy of this form for each publication and each other output</u> for the "top 2" outputs that you selected for each project in Part A above, and provide the following information. Please make your responses brief, but as specific and quantitative as possible. If you consider the criterion not to be applicable to your output, please explain. (Please note that an electronic copy of the questionnaire was included in the email version of this package.)

Name of Output: _____

B1. Technical Quality of Output

In the space below, please describe examples of the technical quality of your output, such as:

- The particular approach or methodology used in developing your output
- Relevant peer recognition such as peer reviews or evaluations, peer endorsements, invitations to present at professional forums or conferences, invitations to present testimony, receipt of awards or honors, etc.
- Receipt of a patent, FDA approval, or use of your output in standards development
- Evidence of the usability and accessibility of the output

B2. Advancement of Knowledge

Please use the space below to describe how this output has advanced knowledge. To structure your response, include points such as:

- What the importance of your original question or issue was
- How the output has advanced knowledge in arenas, such as:
    - making discoveries
    - providing new information
    - establishing theories, measures, and methods
    - closing gaps in the knowledge base
    - developing new interventions, products, technology, and environmental adaptations

B3. Potential Impact

In the space below, please briefly describe evidence of your outputs' potential (or actual) impact on the following audiences, as relevant to your output:

- Science (e.g., new areas of inquiry, methodology, etc.)
- People with disabilities: health, quality of life, participation
- Provider practice
- Health and social systems
- Social and health policy
- Private sector/commercialization
- Capacity building in the field of rehabilitation and disability research and development (e.g., scientists, graduate students, etc.)
- Other

Include information about how this potential impact was tested, and what the results were.

B4. Dissemination of Outputs

In the space below please provide evidence of your dissemination efforts for this output. Describe this for publications if you have made any effort beyond those of the sponsor of the publication (journal, book, proceedings, etc.). Please include important aspects of dissemination such as:

- Stage and scope (e.g., local, regional, national) of dissemination
- Dissemination activities
- Identification and tailoring of materials for reaching different audience/user types
- Collaboration with audience/users in identifying content and medium needs/preferences
- Delivery of information through multiple media types and sources for optimal reach and accessibility
- Evaluation of your dissemination efforts and impacts

Part C. Grant-level Questions

Please respond to these final questions for your overall grant, not by each output specifically as in Section B.

C1. In the space below please describe what types of planning, project management, and budgetary processes were used to promote high quality outputs. In your statement consider the following types of questions:

- Which processes were useful and how? How could they be improved?
- Did you dedicate funds for quality assurance activities?
- How did you track progress and spending against your original plans for the grant?
- If grants or projects were jointly funded by NIDRR and other extramural or intramural sources, how did you ensure that NIDRR resources were used exclusively for NIDRR-funded activities?

C2. Have the results of the research and development outputs from this grant, or prior NIDRR grants, been used to inform the development of new grant applications or other kinds of projects?

No ____
Yes____

If yes, please use the space below to briefly describe what new grant applications, other projects, funding opportunities, or collaborations have emerged.

C3. Please share any perspectives you may have about how NIDRR's key processes (e.g., priority setting, peer review, and/or grants management) influence results, such as successful grants and high quality outputs.

C-4 COMMITTEE MEMBER REVIEW PROCEDURES FOR SUMMATIVE EVALUATION
(Revised 11-08-10)

A. Review Subgroups: Each subgroup that will be reviewing outputs will be composed of five Committee members. For each output one committee member will be assigned as the primary reviewer; the remaining four committee members will be secondary reviewers.

B. Output Rating Procedures:

1. All reviewers will independently rate outputs using the following quality criteria (Dimensions of these criteria are shown on the attached rating sheet.):

- Technical quality of output
- Advancement of knowledge or the field (research, practice, or policy)
- Likely impact
- Dissemination

The following scale will be used for rating the outputs:

Poor Quality			Good Quality			Excellent Quality
1	2	3	4	5	6	7

2. The rating will be based on review of hard copy and electronic materials (i.e., articles/descriptive information about output and questionnaire responses) prior to the subcommittee meeting.

3. The grantee's final summary APR, and a list of all outputs reported over the course of the project, is provided for contextual purposes. The APR also will be used to inform an overall, qualitative grant-level assessment.

4. Multiple outputs of one grant will generally be rated independently of each other. However, in some cases outputs may be rated as a pair with one score applied. This could occur when one output is a derivative or different expression of another output, and when the PI responses to criterion questions are basically the same. Examples of these include:

- A manual describing a device (1) and a patent of the device (2)
- A publication describing how a new technology for assessing a condition can be applied in disability rehabilitation (1) and a description of the technology itself (2)
- A software application (e.g., map reader for persons with visual impairments) (1) and web-based method for individualizing the software for users (2)

5. The meeting will be structured as follows.

- The primary reviewer will open discussion of each output by presenting a brief summary of the output and then his/her rationale for rating each relevant criterion (up to four) plus the overall score,

- Secondary reviewers will then present their ratings for each output and a brief rationale.
- Using the same criteria, the subgroup will then develop consensus group ratings for each output. Discussion will be facilitated by the subgroup chair. If there is a subgroup member with a significantly divergent view, his/her score and rationale will be captured separately.
- Staff will document discussion points that lead to the consensus group ratings and will record the subgroup's rationale for each criterion, the overall rating, and the grant performance rating in a brief narrative.
- At the end of the review of each output, the individual subgroup members' rating sheets will be gathered.

C. Grant Assessment

Once all outputs of an individual grant are reviewed, the subgroup will consider and rate the grant's overall performance. The outputs reviewed had been identified by the grant's Principal Investigator as the "top" two outputs per project, which best reflected the grant's achievements. Taking into consideration this designation, the consensus group ratings of the entire set of outputs, and the grant's overall purpose and objectives (using the grant's APR), the subgroup will assign a grant performance rating using the same 7-point scale. These grant-level ratings and their rationale will also be documented by staff.

C-5 COMMITTEE MEMBER RATING SHEET OF THE QUALITY OF OUTPUTS

Grantee ID: _____

Date of Review: _____

Output to be Reviewed:

To be completed by NRC staff

Output Title: _____

Research Output: _____ Development Output: _____

Type and Subtype of Output (marked below):

Type of Publication		**Type of Tool, Measure, or Intervention Protocol**		**Type of Technology Product or Device**		**Type of Informational Product**	
1. abstract		1. checklist		1. industry standards/guidelines		1. training manuals/curricula	
2. book		2. survey or interview schedule		2. software or netware		2. fact sheets	
3. book chapter		3. diagnostic or assessment instrument		3. invention		3. newsletters	
4. journal article		4. outcome measure		4. patent, license, or disclosures		4. audiovisual materials	
5. proceedings		5. intervention protocol or program		5. working prototype		5. marketing tools	
6. technical		6. statistical technique		6. product evaluated or field tested		6. educational aids	
7. web journal		7. database		7. product transferred to industry for potential commercialization		7. websites or other internet sites	
8. other		8. other		8. product in marketplace		8. other	

Quality Criteria and Dimensions

For each criterion provide one rating using the scale below:

Poor			Good			Excellent
1	2	3	4	5	6	7

Criteria and Dimensions	Score
Technical Quality of Output • Applying standards of science and technology • Appropriate methodology (quantitative or qualitative design and statistics) • Accessibility, usability, etc. **Score Rationale:**	
Advancement of Knowledge or the Field (research, practice, or policy as relevant) • Science: Establishment of methods, tools, theory • New information • Closing an identified gap • New technology • Innovative or novel **Score Rationale:**	
Likely or Demonstrated Impact On: • Science (impact factor, citations) • Consumers (people with disabilities: health, quality of life, participation) • Provider practice • Health and social system • Social and health policy • Private sector/commercialization • Other **Score Rationale:**	
Dissemination • Identification and tailoring of materials for reaching different audience/user types • Collaboration with audience/users in identifying content and medium needs/preferences • Delivery of information through multiple media types and sources for optimal reach and accessibility • Evaluation of dissemination efforts and impacts • Commercialization/patenting of devices, if applicable **Score Rationale:**	
Overall Score **Score Rationale:**	