

Supporting Statement A for

The SSA-NIH Collaboration to Improve the
Disability Determination Process: Validation of
IRT-CAT Tools
NIH/CC/RMD

Date 11/30/11

Name: Meghan Gleason
Address: 10 Center Drive rm 1-2420 NE
Telephone: 301.443.9085
Fax:
Email: Meghan.gleason@nih.gov

Table of contents

A. JUSTIFICATION.....

A.1 CIRCUMSTANCES MAKING THE COLLECTION OF INFORMATION NECESSARY.....

A.2. PURPOSE AND USE OF THE INFORMATION COLLECTION.....

A.3 USE OF INFORMATION TECHNOLOGY AND BURDEN REDUCTION.....

A.4 EFFORTS TO IDENTIFY DUPLICATION AND USE OF SIMILAR INFORMATION.....

A.5 IMPACT ON SMALL BUSINESSES OR OTHER SMALL ENTITIES.....

A.6 CONSEQUENCES OF COLLECTING THE INFORMATION LESS FREQUENTLY.....

A.7 SPECIAL CIRCUMSTANCES RELATING TO THE GUIDELINES OF 5 CFR 1320.5.....

A.8 COMMENTS IN RESPONSE TO THE FEDERAL REGISTER NOTICE AND EFFORTS TO
CONSULT OUTSIDE AGENCY.....

A.9 EXPLANATION OF ANY PAYMENT OF GIFT TO RESPONDENTS.....

A.10 ASSURANCE OF CONFIDENTIALITY PROVIDED TO RESPONDENTS.....

A.11 JUSTIFICATION FOR SENSITIVE QUESTIONS.....

A.12 ESTIMATES OF HOUR BURDEN INCLUDING ANNUALIZED HOURLY COSTS.....

A.13 ESTIMATE OF OTHER TOTAL ANNUAL COST BURDEN TO RESPONDENTS OR
RECORD
KEEPERS.....

A.14 ANNUALIZED COST TO THE FEDERAL GOVERNMENT.....

A.15 EXPLANATION FOR PROGRAM CHANGES OR ADJUSTMENTS.....

A.16 PLANS FOR TABULATION AND PUBLICATION AND PROJECT TIME SCHEDULE.....

A.17 REASON(S) DISPLAY OF OMB EXPIRATION DATE IS INAPPROPRIATE.....

A.18 EXCEPTIONS TO CERTIFICATION FOR PAPERWORK REDUCTION ACT SUBMISSIONS.....

LIST OF ATTACHMENTS:

Attachment 1 – Socio-demographic Information Collected

Attachment 2 – Screener Questions for Participants

Attachment 3 – Security Implementation Plan Overview

Attachment 4– Expert Professionals Consulted

Attachment 5 – IRB Approval

Attachment 6 – NIH Institutional Approval

A.1 Circumstances Making the Collection of Information Necessary

The Rehabilitation Medicine Department (RMD) at the NIH Clinical Center is currently working with the Social Security Administration (SSA) through a multiyear, inter-agency agreement (IAA). The IAA, established in 2008, provides support for the RMD to conduct research focused on improving the SSA's disability determination process. This research includes the feasibility of comprehensively examining function through development of Computer Adaptive Tests (CAT) that could improve the SSA disability determination process.

The Epidemiology and Biostatistics section in RMD will be collecting information through a contractor (Boston University- Health and Disability Research Institute (BU-HDR)) and subcontractor for validation of the CAT tool which is being developed to assist in the SSA disability determination process. The utilization of CAT technology could potentially allow the SSA to collect more relevant and precise data about human functioning in a faster, more efficient fashion.

The NIH/RMD awarded an initial contract to the Boston University Health and Disability Research Institute (BU-HDRI) in 2009 to evaluate the feasibility of integrating a promising new testing method into the SSA's data collection processes. This method uses Computer Adaptive Testing (CAT) coupled with Item Response Theory (IRT) to precisely measure outcomes across the full continuum of human functioning. In order to understand distinct factors influencing work, individual capabilities as well as workplace demands and critical features of the workplace environment must be captured. The contract with Boston University encompasses CAT development to capture the —person side of this interaction, in other words, the assessment of individual capabilities.

The development of CAT tools is a sequentially dependent process. Therefore, each step of CAT tool development proceeds in an ordered fashion; one step must be completed before advancing to the next step. The first step of the process is item pool development. This step encompasses working with content experts, examining literature and reviewing other models/taxonomies to develop item pool content and structure. The next step is to calibrate the items of each pool. Statistical analyses are conducted on data collected from samples of persons similar to the intended audience for the instrument. The objective is to assess the psychometric properties of the items in the pool. The final step of developing CAT tools is to validate and bookmark the instrument, necessary to demonstrate defensibility and to denote cut-points that may aid in disability evaluation decision-making. While the initial contract will develop multiple CAT tool instruments, the content of each instrument is unique and development of each instrument must follow the sequential process.

The first year of BU's contract focused on the identification of functional domains appropriate for CAT instrument development, relevant to SSA's need to determine work disability. The physical demands and interpersonal interactions and relationships were selected as the initial domains for development. The selection was motivated by prior work on mobility CAT instruments that could be tailored for SSA's needs; and, by SSA's

desire for improved approaches to evaluate claimants with cognitive and mental health conditions.

The initial project design projected the development of a single CAT tool for each distinct domain of health considered important to determine work disability. However, increased understanding about the SSA disability determination process revealed that one CAT tool for each domain of health would be insufficient. SSA collects information from claimants and medical providers to inform determination decisions. It is unlikely that providers will be able to answer detailed questions about activities that their patients perform in the work place since providers do not assess patients in these environments and generally do not assess work-related activities. Appreciating the challenges associated with gathering information from health care providers about patients' functioning, it was determined that distinct applicant and provider CAT instruments would be needed. Thus, development of a CAT tool composed of more general provider questions about a patient's performance of functional activities, in addition to a CAT tool with more specific applicant questions about performance is necessary. This approach will better support the validity of the information collected from each of the user populations. For each domain of health, two CAT tools would need to be developed instead of one. Claimant and provider CAT instruments for the physical demands and interpersonal interaction and relationship activity domains are in progress.

Development of the item pools for the two functional domains is an iterative process influenced by the literature, existing instruments, content experts, focus groups, and cognitive testing. BU developed detailed schematics of the content models for both domains and operationally defined terminology to facilitate clarity and enhance precision of the sub-domains encompassed within each model. Upon conclusion of the contract period, BU provided full item pools for the physical demands and interpersonal interactions and relationships domains. They also submitted a detailed design for a calibration study necessary to test the "fit" of the items included in each domain and detect the presence of gaps in item content. BU completed development of the software needed to collect the calibration study data. NIH approval of the calibration study may be found in Attachment 6.

A calibration study is a field study of item content and structure conducted with samples of respondents representing the intended users of the CAT instruments. Item pool development and subsequent item calibration are unique for each CAT tool and for the target population for which they are developed. Sample sizes for scientifically defensible CAT calibration projects are in the range of 500-700 respondents. This sample size is needed to support a series of confirmatory factor analyses and to perform statistical modeling. Inadequate sample size may lead to inaccurate and unstable statistical outcomes. To better understand the breadth of medical providers participating in the care of SSA claimants and to capture the varied perspectives across medical disciplines, up to two providers for each claimant participating in the study will be contacted. For each claimant, contact information for a primary provider and a supplementary provider will be requested. It is estimated a substantial number of calls to potential participants will be necessary in order to recruit 500 claimant-provider dyads (claimant-provider triads if two

of the claimant's providers participate) for a calibration study in each of the two domains, mobility and interpersonal interactions and relationships.

The SSA-NIH-BU/Westat team collaborated to establish the calibration design and sampling strategy. SSA's Office of Data Analysis extracted from SSA automated records claims submitted within the last three months who met the following selection criteria: claimant name, address, phone number, impairment allegations on the form SSA-3368BK. From this dataset, Westat assigned a geographic variable to the data to classify claimants into urban or rural categories. The sample was then stratified by urban/rural status across the 10 national SSA office regions. A randomly selected subsample was then drawn in each of the two study groups (domains), the physical domain and the interpersonal domain. Each study group contained 5,000 beneficiaries. Westat contacted and notified the beneficiaries in each domain about the study. Data were collected from 1,015 claimants in the interpersonal domain, 1,017 claimants in the physical domain; and, a total of 110 providers. The claimant calibration data are presently being analyzed to develop optimal model solutions necessary to complete the IRT/CAT software.

In addition to collecting data from SSA claimants, data from a normative sample was collected and currently under analysis. The normative sample data allows the research team to expand the breadth of each scale developed compared to use of claimant data alone. This will reduce ceiling effects and broaden the overall applicability of each CAT scale developed. Secondly, having calibration data from a normative sample of adults in the country provides a useful reference population against which SSA claimants can be compared. This allows SSA to better characterize their population of claimants over time.

The normative national sample was obtained using sample matching; a methodology pioneered by YouGov Polimetrix, Inc. (YGP; Palo Alto, CA) whereby samples representative of a study-appropriate target population can be constructed from large but unrepresentative pools of opt-in survey respondents. The enumeration of the target population would in traditional sampling be known as the sampling frame and would serve as the source from which the sample would be drawn. This is not the case in sample matching, which instead proceeds in two-stages. First, a random sample is drawn from the enumeration of the target population. A simple random sample (SRS) could be drawn; but in practice, the efficiency of the procedure can be improved by using stratified sampling. YGP typically stratifies on race, gender and age, and then draws a SRS from each of the mutually exclusive and exhaustive groups formed by the simultaneous cross-classification of the population on these three attributes. The SRS from each category is combined to form the stratified target sample. If the number of respondents selected in each stratum is proportional to their frequency in the target population, then the sample is self-representing.

Conventionally, one would then attempt to contact the respondents in the target sample. However, there is no economical way of reaching most members of the target sample, as they have not provided their email addresses and many do not have a listed phone number, and those that do, may not agree to be interviewed. Instead, for each member of the target sample, YGP will select one or more matching members from their pool of opt-

in respondents. This pool has been recruited by a variety of means and currently numbers approximately 1.5 million. Of course, data drawn from this pool would not be representative of any particular population; individuals who opt-in for taking web surveys have different demographics than either the population of all internet users or the population of all adults. Rather, the matching methodology is required to produce usable samples for individual studies. Matching is done on a large set of variables available in both the population enumeration database and the opt-in panel. The purpose of the matching is to find an available respondent who is as similar as possible to the selected member of the target sample. YGP employs a proximity matching method whereby a distance function is computed for each attribute to define the degree of “closeness” between each individual in the target sample (x) and those in the opt-in survey panel (y). Typically, the distance function is the simple absolute value of the difference, $|x-y|$, and the overall distance between a member of the target sample and a member of the panel is a sum of the distance functions for each attribute being used in the matching. The distance functions can be weighted and then summed if particular variables are thought to be more important for a given study. For this study, matching was done on gender, racial/ethnic background, age, education and employment status, weighted equally.

YGP adjusts for anticipated non-response by selecting multiple best matches in the opt-in panel for each member of the target sample. The number of matches is determined by using a hazard model to estimate the probability that an opt-in panelist will respond by the end of the data collection period, and increasing the number of panelists matched to the member of the target sample until that response probability is ≥ 1 . Although internet use was initially concentrated in the more affluent and better-educated segments of the population, this “digital divide” has been substantially reduced such that nearly three-quarters of the adult population now have access to the internet either at home, work or school.

YGP used web survey administration to deliver the same item pool surveys used with SSA claimants to 1000 matched normative adults in each domain, allowing analysis of both data sets together. Once analyses and the IRT software are complete, the validity of the IRT/CATs must take place.

To validate the CAT instruments that have been developed, the contractor will administer both the BU-HDR CAT and established legacy instruments in a small sample of adults who report their current employment status as “permanently disabled”. Individuals will complete the CAT tools for the functional domains of Physical Demands and Interpersonal Interactions along with established legacy instruments. For the domain of physical function, individuals will complete the BU-HDR CAT; the PROMIS Item Bank v 1.0-Physical Functioning © PROMIS Health Organization and PROMIS Cooperative Group; and, The Short Form (36) Health Survey™ (SF-36). For the domain of interpersonal interactions, individuals will complete the BU-HDR CAT, the SF-36 and the BASIS-24© (Behavior and Symptom Identification Scale). Data collected will be used to validate the BU-HDR CAT tools. Without this information, completion of the BU-HDR CAT tools will not be possible.

A broad timeline for CAT tool development is as follows:

Year 1	Year 2	Year 3
Develop item banks for disability applicants and medical providers in areas of functioning	Calibration tests of item banks developed	Building CAT software, validation and bookmarking of CAT instruments in functional domains

Disability, in this circumstance—SSA’s perspective of work disability, is the interaction between the functioning of the whole person and environmental demand. The assessment of functioning provides SSA a mechanism to integrate contemporary perspectives of disablement into disability program processes. The use of IRT/CAT assessments may allow SSA to capture functional information in a more precise, efficient and comprehensive manner. This may improve the uniformity of decision-making and potentially reduce program costs by informing decision-making earlier in the evaluation process.

This initiative is authorized by the Public Health Service Act, Title 42> Chapter 6A> Subchapter III>Part B> § 284k. This states that the Director of National Institutes of Health shall undertake activities to support and expand the involvement of the National Institutes of Health in clinical research.

A.2 Purpose and Use of the Information Collection

The information for the proposed data collection will be collected by the NIH Clinical Center through a contract with Boston University and sub-contract with YouGovPolimetrix (Polimetrix), a survey research firm based in Palo Alto, CA. This information will be used to validate the BU-HDR CAT instruments. The proposed information collection will support psychometric testing.

In order to accomplish this, NIH proposes to recruit 1,000 adults through the Polimetrix survey center. Polimetrix recruits for studies using an opt-in panel of 1.5 million U.S residents who have agreed to participate in Polimetrix’s Web surveys. Panel members are recruited by a number of methods to help ensure diversity in the panel population. Recruiting methods include Web advertising campaigns (both text and banners), permission-based email campaigns, partner sponsored solicitations (*e.g.*, Rock the Vote and Cox Communications), telephone-to-Web recruitment, and mail-to-Web recruitment.

By utilizing different modes of recruitment continuously over time, this ensures that hard-to-reach populations will be adequately represented in survey samples. Participants are not paid to join the PollingPoint panel, but do receive modest incentives through a loyalty program to take individual surveys (see question A.9 for detailed explanation of incentives).

Polimetrix tracks employment status within their active participant pools. They currently have about 5,600 "permanently disabled" participants which will serve as the sampling frame for this validation work. From this population, Polimetrix will recruit 1,000 participants to answer 70-96 items depending on their primary reason for disability.

The participants will be matched on age, gender, race, and education with the participants that participated in the BU-HDR CAT tool's calibration sample. Study participants will provide basic sociodemographic information (Attachment 1) as well as answer a screener question (Attachment 2) to help determine if the reason for their "permanently disabled" employment status is the result of a primary physical or mental health impairment. This information will be used to match each potential respondent to the appropriate CAT content domain (i.e., Physical Demands or Interpersonal Interactions).

Study participants who indicate that their primary reason for not being able to work is the result of a primary physical impairment will be asked to complete the BU-HDR CAT for Physical Demands, the SF-36, and the PROMIS Physical Functioning CAT. The SF-36 is a widely used multi-purpose, short-form survey with 36 questions that measure physical and mental health. It has been broadly tested in general and disease specific populations. The SF-36 consists of eight scaled scores: vitality, physical functioning, bodily pain, general health, physical role functioning, emotional role functioning, social role functioning, and mental health. The SF-36 was selected as a legacy instrument because of its extensive history and use in research and the content coverage it provides. The PROMIS Physical Function item pool consists of items covering activities of daily living, lower extremity, and central body functions. PROMIS utilizes rigorous methodology for developing its measures and testing their validity. This work integrates qualitative and quantitative research and psychometrics. Content and disease experts as well as thousands of patients provided input into the development process. The PROMIS item pools have been tested and validated in clinical and generic populations. The PROMIS physical function CAT was selected because of the content coverage it provides as well the extensive and rigorous testing process the PROMIS initiative utilized. The PROMIS CAT, unlike the BU-HDR CAT, was not developed to assess functioning relative to work disability.

Study participants who indicate that their reason for not being able to work is the result of a primary mental health impairment will be asked to complete the BU-HDR CAT for Interpersonal Interactions, the SF-36 and the Behavior And Symptom Identification Scale (BASIS-24.) The SF-36 is a widely used multi-purpose, short-form survey with 36 questions that measure physical and mental health. It has been broadly tested in general and disease specific populations. The SF-36 consists of eight scaled scores: vitality, physical functioning, bodily pain, general health, physical role functioning, emotional

role functioning, social role functioning, and mental health. The SF-36 was selected as a legacy instrument because of its extensive history and use in research and the content coverage it provides. The BASIS-24 is a leading behavioral health assessment tool. The BASIS-24 underwent extensive field testing as part of a multiyear research and development process. The survey was tested on more than 6,000 participants from racially and ethnically diverse backgrounds receiving inpatient or outpatient treatment for mental health or substance abuse at one of 28 facilities across the U.S. The development of the survey was grounded in Item Response Theory (IRT) methods. The substance use sub-scale (4 questions) will be removed from The BASIS-24 for the purposes of this validation study.

Actual respondent requirements will differ slightly depending on domain of functioning identified (See Table 1).

Table 1. Summary of Survey Content in Two Domains

Domain	BU-HDR CAT Tool	SF-36	PROMIS PF CAT	BASIS-24	Total Per Domain
Physical Demands	24-30 items	36 items	10-20 items	0	70-86 items
Interpersonal Interactions	32-40 items	36 items	0	20 (4 alcohol/drug items removed)	88-96 items
Total	56-70 items	72 items	10-20 items	20 items	

The responses to the BU-HDR CAT tools and the legacy instruments will be used by NIH and its contractor Boston University to examine the psychometric properties of the BU-HDR CAT tools to validate its use compared to legacy instruments

A.3 Use of Information Technology and Burden Reduction

The proposed validation study will collect all data electronically on a computer and individuals will submit responses electronically. All of the Polimetrix panelists have provided their e-mail so that they may receive electronic invitations to participate in surveys. Additionally, with each survey invitation they are reminded of the Polimetrix policy on privacy, the opportunity to immediately opt-out, and of the voluntary nature of each request regardless of the survey sponsor.

Electronic data capture using computer adaptive testing technology is more efficient compared to fixed form assessment instruments and substantially reduces respondent burden.

NIH and its contractor currently has extensive security and privacy agreements in place to support the NIH-SSA IAA (See Attachment 3). Additionally, a Privacy Impact

Assessment (PIA) has been performed for work related to this project at the NIH. However, a specific PIA will be performed at Boston University for the specific data related to this information collection.

The estimated respondent burden is 0.5 hour of human element and no financial cost to the participants is anticipated.

A.4 Efforts to Identify Duplication and Use of Similar Information

There is no duplication of effort or similar information available for use. This is a new CAT tool developed specifically for the SSA disability programs and requires validation prior to pilot testing. The actual validation data to be collected do not currently exist. The data will be unique to the instrument currently being developed and will feed back into the psychometric evaluation of the assessment instrument. Data are necessary to provide a basis and a context for validating outcomes of the new instrument with existing assessment instruments considered the “gold standard.” With a sample size of 500 subjects per domain we anticipate sufficient variation in duration of disability and that we will be able to select adequate numbers of those with relatively recent disabilities post hoc for our analyses.

A.5 Impact on Small Businesses or Other Small Entities

No small businesses will be involved in this data collection.

A.6 Consequences of Collecting the Information Less Frequently

Information will only be collected once from each participant and study participants will not be re-contacted.

A.7 Special Circumstances Relating to the Guidelines of 5 CFR 1320.5

This project fully complies with 5 CFR 1320.5

The data collection in this project will support the development of a measurement instrument (i.e., methodological development) to assess functioning with respect to work disability. Results should not be generalized to other populations. Respondent burden is estimated to be 0.50 hours per respondent.

A.8 Comments in Response to the Federal Register Notice and Efforts to Consult

Outside Agency

The 60-day Federal Register notice was published on 9/8/2011: Federal Register Volume 76, Number 174. No comments or questions were submitted.

Consultation and coordination has been sought throughout the BU-HDR CAT development process. CAT tool developers examined items from existing NIH tools

(such as those developed for PROMIS and Neuro-QOL) for potential inclusion in the computer adaptive tests being developed by BU. PROMIS aims to use computer adaptive testing methodology to develop ways to measure patient-reported symptoms, such as pain and fatigue, and aspects of health-related quality of life across a wide variety of chronic diseases and conditions. The Neuro-QOL is a 5-year, multi-site project funded by the National Institute of Neurological Disorders and Stroke (NINDS) which is intended to develop assessments that address dimensions of health-related quality of life that are universal to adults and children with chronic neurological disorders. Neuro-QOL is also based on patient-reported outcomes and uses CAT methods to assess pain, fatigue, emotional distress, physical function, and social function. Since both PROMIS and Neuro-QOL use CAT methods to assess patient-reported outcomes, items from these assessments could be selected for inclusion in both the physical demands and interpersonal interaction item banks being developed by BU. In fact, 57 of the 361 items for the interpersonal interactions item pool for claimants were drawn from PROMIS and Neuro-QOL items, representing 16% of the total number of items. In addition, 31 of the 128 items for the physical demands item pool for claimants were drawn from PROMIS and Neuro-QOL items, representing 24% of the item pool.

As BU began CAT development for the physical demands and interpersonal interaction domains, they evaluated existing conceptual frameworks in order to develop the structure for each domain. Existing conceptual frameworks were consulted including those for PROMIS, Neuro-QOL, and the NIH Toolbox (assessment of neurological and behavioral function) projects. While these existing frameworks were developed for specific populations and a different purpose, they were critical in informing domain structure for the BU CATs. As BU continued CAT development for these two domains, the PROMIS and Neuro-QOL assessments were also used to guide preliminary development of domain content, focusing on content coverage within each major domain and related sub-domains. In all of these ways, the NIH PROMIS, Neuro-QOL and Toolbox initiatives have heavily informed development of the two initial BU-HDR CATs.

Consultation was also sought by BU CAT development experts on which legacy instruments should be used to validate the CAT tool. (See attachment 4)

A.9 Explanation of Any Payment of Gift to Respondents

Polimetrix's goal is to provide a small thank you for a respondent's time, but not an incentive that might make survey response a financial transaction. The average survey incentive of 500 points cashes out at 50 cents, although it's not redeemable until respondents reach certain thresholds. Respondents who complete the 3 instruments required for this validation study will earn 1500 points.

A.10 Assurance of Confidentiality Provided to Respondents

With each survey respondents are reminded of the Polimetrix policy on privacy, the opportunity to immediately opt-out, and of the voluntary nature of each request regardless of the survey sponsor.

Responses to questions will remain secure to the extent permitted by law. Respondents will not be identified by name. The link to codes and names will be destroyed after the study completion and after the acceptance for publication, if appropriate. Any information respondents provide will be available only to research staff. All information respondents provide in this study will be only for research purposes and their name will not be used in any publication that may be written from this research. This research project will be conducted in compliance with all applicable state and federal laws.

When BU receives the raw data back from Polimetrix, it will only contain respondent ID numbers and no other personal identifying information (such as name, street address, e-mail address, phone number, or social security number).

This project has been approved by the Boston University Institutional Review Board. A copy of the IRB project approval is included as Attachment 5. This research is conducted under the NIH system of records listing notice (SORN), 09-25-0200; system name: Clinical, Basic and Population-based Research Studies of the National Institutes of Health (NIH), HHS/NIH/OD.

A.11 Justification for Sensitive Questions

Questions included in the CAT instrument for interpersonal interactions as well as the BASIS=24 © (Behavioral And Symptom Identification Scale) are designed to assess a respondents functioning in interpersonal domains. This includes sensitive questions that can be regarded as “psychological problems” including questions about feelings towards others, and mood swings. These questions are essential to validation as they are included in the instruments, including the legacy instruments which are widely used and will be compared to the BU-HDR CAT instrument. It would be impossible to complete CAT validation without including these questions. The substance use sub-scale will be removed from The BASIS-24 for the purposes of this validation study.

While Polimetrix retains personally identifiable information (PII) for individuals who choose to participate in their surveys (name, address) the NIH along with its contractor, Boston University, will not be provided with that information. Boston University will be provided with de-identified data that only includes demographic information including: age, race, gender, marital status, education, and zip code.

Respondent consent will be obtained by Polimetrix. Polimetrix will retain responsibility and oversight of the consent process. Since all respondents will take the survey online

they will review and click through a consent text form prior to being able to start the survey. A waiver of documentation of informed consent has been obtained from BU's IRB.

A.12 Estimates of Hour Burden Including Annualized Hourly Costs

It is estimated 1,000 individuals will participate in the validation study. They will be asked to respond only once and the 0.5 hour of burden is a total estimate of the time to complete the individual assessment tools for one domain. Study participants who indicate that their primary reason for not being able to work is the result of a primary physical impairment will be asked to complete the BU-HDR CAT for Physical Demands, the SF-36, and the PROMIS Physical Functioning CAT. Study participants who indicate that their reason for not being able to work is the result of a primary mental health impairment will be asked to complete the BU-HDR CAT for Interpersonal Interactions, the SF-36 and the Behavior And Symptom Identification Scale (BASIS-24.) Regardless of which domain respondents are selected for; a 0.5 hour of burden is the total estimate of the time to complete the individual assessment tools for one domain. The estimates of hour burden provided below are based on the research experience during a similar item bank development project, which respondents completed similar types of items. With respect to time costs, all individuals completing the instruments will have indicated that they are "permanently disabled" participants. We are therefore assuming, for the purposes of this validation study, that these individuals are currently not employed.

A.12 - 1 ESTIMATES OF HOUR BURDEN				
Type of Respondents	Number of Respondents	Frequency of Response	Average Time per Response	Annual Hour Burden
Patients	1,000	1	0.5	500.00
Totals				500.00

A.12 - 2 ANNUALIZED COST TO RESPONDENTS

Type of Respondents	Number of Respondents	Frequency of Response	Average Time per Respondents	Hourly Wage Rate	Respondent Cost
Patients	1,000	1	0.5	\$0.0	\$0
Totals					\$0

A.13 Estimate of Other Total Annual Cost Burden to Respondents or Record Keepers

There are no costs to respondents beyond time.

A.14 Annualized Cost to the Federal Government

This validation study is being supported through an NIH contract with BU, Contract No. HHSN269201100009I: Applying CAT testing to the SSA Disability Evaluation Process. It is estimated that the cost of subcontracting to YouGovPolimetrix for the data collection portion of the validation will cost \$52,000.00

A.14 - 1 Estimate of Annualized Cost to the Federal Government	
Total Capital and Start up Costs for Polimetrix through contract with BU	\$4200.00
Total Operations and Maintenance for Polimetrix through contract with BU	\$2800.00
Cost of Data Collection	\$44,500.00
Portion of Contract Costs for Boston University Personnel to support information collection and analytic work	\$55,000.00
FTE costs at NIH/CC/RMD	\$2,000.00
Total	\$108,500.00
Estimate of Annualized Costs to the Federal Government:	\$108,500.00

Estimates for Polimetrix costs are based on market research. Contact costs and FTE costs are estimated based on hours of personnel supported required to complete data analysis and project management.

A.15 Explanation for Program Changes or Adjustments

This is a new collection of information

A.16 Plans for Tabulation and Publication and Project Time Schedule

This validation study is part of a larger scientific project focused on improving SSA's disability determination process. Data collection is projected to begin in June 2012 and completed by August 2012. Subsequent data analysis should be completed by January 2013. This work is part of an existing contract with BU. These data will be analyzed and outcomes published as part of the larger project work; however, CAT tool development for the two domains included in this validation is scheduled to be completed by May 2013.

The data analysis will address the following parameters:

- Response burden
- Score precision
- Internal consistency reliability
- Score range (ie., floor or ceiling effects)
- Concurrent validity

To monitor the BU-HDR CAT in real time, we will calculate the standardized log-likelihood statistic (l_z) for polytomous items to test the person fit. The empirical distribution of the log-likelihood statistic is reasonably close to a standardized normal distribution, so we will calculate the percentage of respondents in which l_z exceeded an alpha level of .05.

Response burden will be measured as the average amount of time it takes to complete instrument. A t-test will be used to assess whether the average amount of administration time between the BU-HDR CAT and other measurements is significantly different.

To illustrate the difference in precision in score range across instruments, we will calculate the average Standard Error (SE) along the entire scale continuum across different instruments. We will use the t-test to assess whether the average SE is significantly different between BU-HDR CAT and other measurements at different score ranges.

To examine internal consistency, we will use marginal reliability calculations that are specific to item response theory (IRT) which allow us to compare BU-HDR CAT with other instruments. Marginal reliabilities are similar to Cronbach's alpha coefficient used in classical measurement theory in that it is a measure of how well items within a domain relate to each other.

The percentage of ceiling and flooring will be calculated in each instrument. A chi-square test will be used to test whether the percentages of ceiling or flooring are significant different between BU-HDR CAT and other instruments.

We will analyze the concurrent validity of the BU-HDR CATs and other measures using Pearson correlation coefficients. Specifically, Pearson correlations coefficients will be calculated between scores from the BU-HDR physical function CAT and the SF-36 scale scores, and the PROMIS Physical Function CAT scores; between the BU-HDR interpersonal interactions CAT and the SF-36 scale scores, and the BASIS-24 scale scores.

A.16 - 1 Project Time Schedule	
Activity	Time Schedule
Invitation to Participants from Polimetrix	1 - 2 weeks after OMB approval
Online data collection	0.5 - 2 months after OMB approval
Analyses	4-6 months after OMB approval
Tool Validation Complete	12 months after OMB approval

A.17 Reason(s) Display of OMB Expiration Date is Inappropriate

Data collection will be conducted by web administered surveys; the OMB expiration date will be displayed on the data collection screen.

A.18 Exceptions to Certification for Paperwork Reduction Act Submissions

No exceptions are requested.