IEEE Access
The journal for rapid open access publishing

Received April 11, 2014, accepted May 8, 2014, published July 16, 2014, date of current version July 29, 2014.

Digital Object Identifier 10.1109/ACCESS.2014.2339328

Cognitive Ability-Demand Gap Analysis With Latent Response Models

GAHANGIR HOSSAIN[1,2], (Student Member, IEEE), AND MOHAMMED YEASIN[1], (Senior Member, IEEE)
[1]Department of Electrical and Computer Engineering, University of Memphis, TN 38152, USA
[2]Chittagong University of Engineering and Technology, Chittagong 4349, Bangladesh

Corresponding author: G. Hossain (gahangir@gmail.com)

This work was supported by the National Science Foundation under Grant NSF-IIS-0746790.

ABSTRACT A better understanding of human cognitive ability-demand gap (ADG) is critical in designing assistive technology solution that is accurate and adaptive over a wide range of human-agent interaction. The main goal is to design systems that can adapt with the user's abilities and needs over a range of cognitive tasks. It will also enable the system to provide feedback consistent with the situation. However, the latent structure and relationship between human ability to respond to cognitive task (demand on human by the agent) remains unknown. Robust modeling of cognitive ADG will be a paradigm shift from the current trends in assistive technology design. The key idea is to estimate the gap, based on human-agent cognitive task interaction. In particular, latent response model was adopted to quantify the gap. First, we used one parameter Rasch model and extended Rasch model (rating scale model, partial credit model) with dichotomous and polytomous responses, respectively. Residues between expected and observed ability scores were considered as gap parameter in case of dichotomous response. In extended Rasch modeling, response latitudes are considered as an indicator of the gap. Additionally, we performed model fitting, standard error measurement, kernel density estimation, and differential item functioning to test the suitability of Rasch model. Empirical analyses on a number of data set show that proposed analytical method can model the cognitive ADG from dichotomous and polytomous responses. In dichotomous case, the model better fits for mixed responses (combination of easy, medium, and hard) data set rather than monotonic (e.g., only easy) data. Results show that Rasch model can be reliably used to estimate cognitive gap with different cognitive task types.

INDEX TERMS Cognitive engineering, ergonomics, human computer interaction, human factors, cognitive modeling, cyber physical systems.

I. INTRODUCTION

Blind Ambition is an ongoing innovative research project at the Computer Vision Perception and Image Analysis (CVPIA) laboratory at the University of Memphis. The main aim is to develop accurate, adaptive, affordable, effective and portable assistive technology solutions for the people who are blind or visual impaired. The key idea is to keep the design simple so that the user can interact with the system effectively with minimal cognitive effort. A number of products and wide varieties of services have been developed –reconfigured mobile android phone application (RMAP [18] iMAP [19], EmoAssist [20], FEPS [21], [22] and E-Glass.

Designing an adaptive assistive technology solutions require an understanding of user's need and ability to use the system to perform the task with minimal cognitive effort. Traditional designs focus on providing the functionalities without considering the adaptive behaviorism - a type of behavior that is used to adjust to another type of behavior or situation [1], [2]. Using mobile devices in such technology solicitation is a challenging task [7], [8]. The research presented in this paper is a step towards bridging the ability-demand gaps (ADGs) in developing assistive solutions.

A. ABILITY-DEMAND GAP (ADG)

The discrepancy between user's ability and cognitive demand on the user by technology solutions remains the source of incoherence in human-machine (agent) interaction. This discrepancy is known as (dis)ability or ability-demand gap (ADG) [1], [3], [9]. Modeling and quantification of the ADG is critical in designing the next generation human-agent

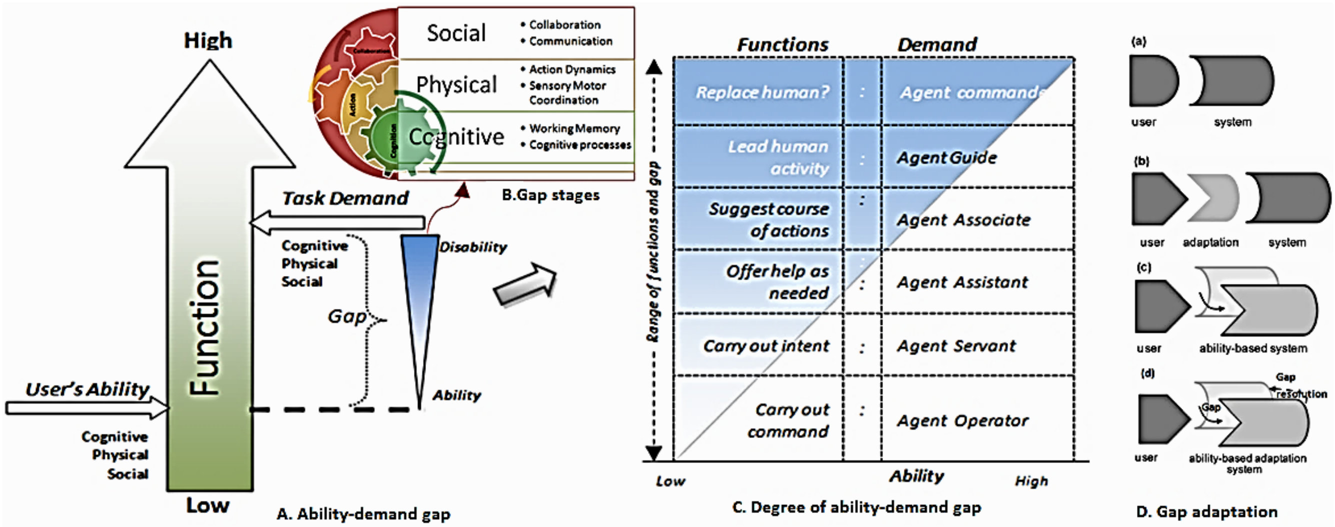


FIGURE 1. A conceptual illustration of interaction gap (cognitive, physical, collaborative) - (A) Ability-demand gap - the gap model, (B) Gap stages - gap amplification with a layer of hierarchy (C) Degree of gaps - range of gaps in human-agent (system) interaction, and (D) Gap adaptation - in human-system (agent) interaction.

interaction system and assistive technology solutions. It is easy to note that a system capable of understanding and responding to the gap will help in error resolution, minimizes incoherent interaction, and facilitates meaningful and personalized feedback and also will improve users interaction experiences.

The Fig. 1 illustrates a number of issues related to the gap and system (agent's) performance under different roles. The Part A depicts the cognitive ability-demand gap from a utilitarian point of view; part B shows different levels of possible gaps and their associations with cognitive, physical and social (collaborative) attributes; part C presents agents different roles as a function of increasing degree of ADGs and part D illustrates how the gap can be used to resolve issues in human-agent (computer) interaction.

Continuous monitoring of gaps will enable designing a multi-layer feedback system that is adaptive with the complexity of the cognitive task and consistent with user's need. This research considers only the cognitive ADG in the context of human-agent interaction. A range of other types of gap can be considered depending on the agent's role, and cognitive functions [4]–[6].

This study considers ADG as the interplay of human ability and task's demand, like (but, not similar) to Yin-Yang philosophy. Details can be found from authors' earlier work [49]. Human expected demand sometime corresponds to their ability, so the ability and demand are cluttered together as like the fig.2A. The latent properties of the interplay are illustrated in fig.2B. Subject has his ability (θ) parameter, and the task it's discrimination (α) parameter, both are aligned with task difficulty (δ). Notably, ability-demand gap resolution mitigates shared understanding between the user and assistive technology tools (e.g. Mobile application).

Past observations applied item response model in the ability estimation in psychometric assessment [31]–[35], clinical usability assessment [31], [32], [36], cyber physical systems [10] and political science [29]. The ability-demand gap paradigm can be used with assistive intelligent tutoring system [17], multimodal user interface in crisis management [23], brain computer interfacing research [24], assistive technology design [15] and collaborative sense-making [12], [13]. This study proposes item response modeling approach in human computer interaction with ability - demand gap assessment. With cognitive ADG identification being the primary objective of the paper, Section II discusses the literatures related to ability-demand gaps computation, modeling, and relation to human computer interaction and cyber physical systems.

Next, the gap computation is examined with two datasets which is explained in Section III. Section IV shows empirical analysis results with dichotomous and polytomous data, item-person maps and fit computation. Finally Section V concludes the paper.

II. ABILITY-DEMAND GAP ANALYSIS

In this section, we review different computational approaches and their application in estimating the anility-demand gap (ADG) from human-agent interaction. We explain how different parameters of the latent response models are related to the gap.

A. ADG COMPUTATION

Disability can be viewed as the difference between the cost demanded by environment and individual's ability [1], [2]. In new technology adaptation, a user starts with ADGs that are reduced with an increase of skill and experiences. Users with physical or sensory disability will always have some

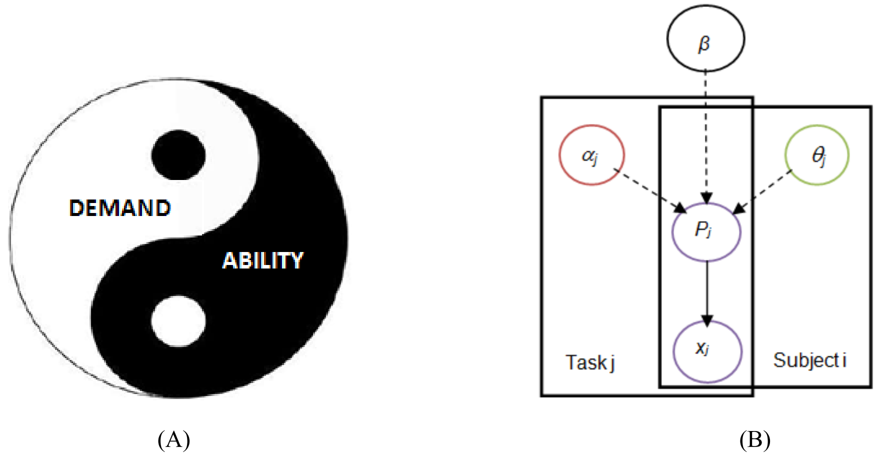


FIGURE 2. Ability-Demand gap Illustration, (A) Interplay of ability and demand [49], (B) latent diagram with subject and task parameters.

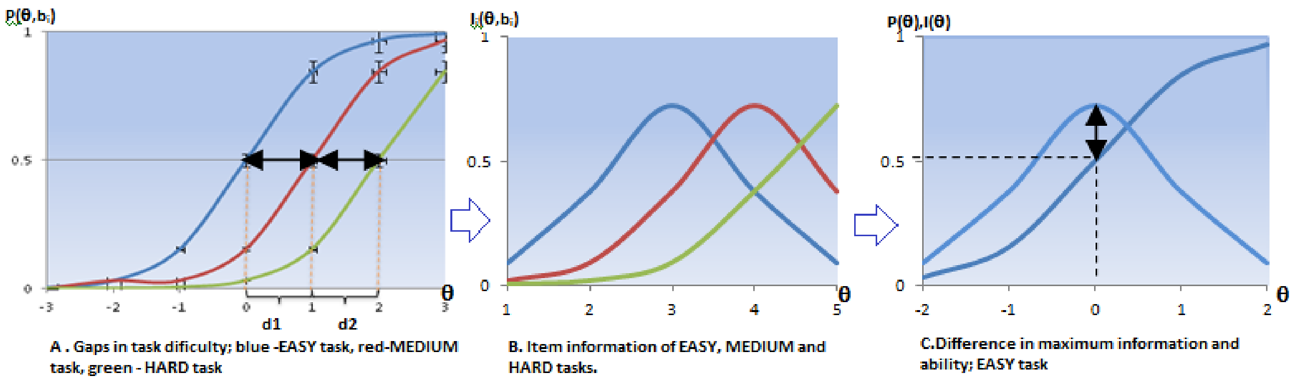


FIGURE 3. Residual computation of ability-demand gap.

gaps (in their ability) to be considered with holistic design. The ADG can be formalized as a latent response analysis as illustrated in Fig. 2B, with subject's ability vs. task demand with task difficulty (β), task discrimination (α), subject's ability (θ), performance (x) and the probability of success ($P_{j)}$. Intuitively, the cognitive demand on the user can be written as [25]:

$$D = f\,(P_d, Ph_d,\ S_d) + D_A \tag{1}$$

Where, D is the overall demand, P_d is the physical demand; P_{hd} is the psychological demand; S_d is the sociological demand, and D_A is demand with ignored ability. The function on the right side of the equation is the demand function. Similarly, the ability can be expressed as [25]:

$$A = f\,(P_a, Ph_a,\ S_a) + A_D \tag{2}$$

Where, A is the overall ability, Pa is the physical ability, P_{ha} is the psychological ability; S_a is the sociological ability and A_D is the ability with ignored demand. The function on the right side of the equation is the ability function. The ability demand gap can be formalized as"

$$G = K(D - A) \tag{3}$$

Where, G is the ADG, K can be considered as a normalizing constant (e.g., the power constant 0.74±0.06, 95% confidence limits [28]).

B. DIFFERENCE COMPUTATION

Given the subject's ability A and cognitive task demand D, the ADG can be defined by Coombs theory of difference [28]. If A is greater than D, say, A − D > 0 and the subject make some error. With probability of error the equation can be written as,

$$p\,(A > D|A, D) = (A - D) \tag{4}$$

Alternatively, A is close to D if the absolute difference between them, is less than some threshold, δ.

$$p(|A - D| < \delta|AD\delta) = f(|A - D|, \delta) \tag{5}$$

This distinction are shown graphically by considering the probability of being greater as a function of the distance A -B (Fig 3.c) or the absolute difference between A and B. Ordered difference consider the probability of observing A > D as a function of the difference between A and D. The greater the signed difference, the greater the probability that A will

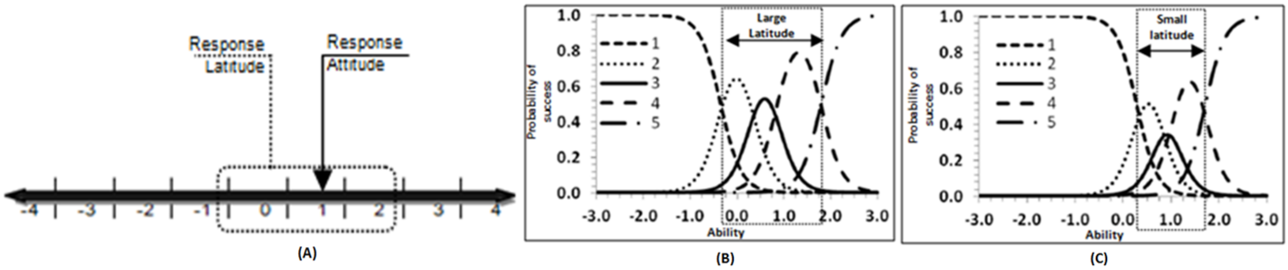


FIGURE 4. **Response latitude (RL) computation - (A) response attitude and latitude, (B) Large latitude, (C) Small latitude.**

be reported as greater than D. The three lines represent three different amounts of sensitivity to distance. The proximity relationship considers the probability of observing A is the same as (close to) D as a function of the difference between A and D. The less the absolute difference, the greater the probability they will be reported as the same. Given a data matrix D with features $d_{ij,}$ we try to find model values m_i and m_j such that some function f when applied to the model values best recreates d_{ij}. For data that are expressed as probabilities of an outcome, the model should provide a rule for comparing multiple scale values that are not necessarily bounded 0−1with output values that are bounded 0−1. That is, we are interested in a mapping function f such that for any values of m_i and m_j

$$0 \leq f(mi, mj) \leq 1 \tag{6}$$

In order to fit it to model, we need to find scale values that minimize some function of the error. Applying f (m_i, m_j) for all values of i and j produces the model matrix M. Let the error matrix E = Difference (f-norm (D), f-norm (M)). Because average error will tend to be zero no matter how badly the model fits; median absolute error or average squared error are typical estimates of the amount of error. A generic estimate of goodness of fit in terms of errors becomes

$$GF = f\,(D, M) \tag{7}$$

Variations on this generic goodness of fit estimate include Ordinary Least Squares Estimates such as

$$GF = (D, M)^2/(D)^2 \tag{8}$$

and measures of median absolute deviation from the median, or many variations on Maximum Likelihood Estimates of $\chi 2$.

C. ITEM RESPONSE MODEL AND GAP COMPUTATION

The item response theory (IRT) model predicts the probability that a certain subject gives a certain response to a certain item. In a single item response setting, let the subject x may only have dichotomous responses (1 = correct, or 0 = incorrect), let P_{ij} as the probability of a correct response, where i refer to the task, and the index j refers to the subject. The function shown on the graph is known as the one-parameter logistic function.

$$\mathrm{P_{ij}}(\theta_j, \mathrm{b_i}) = \frac{1}{1+\mathrm{e}^{-(\theta_j - \mathrm{b_i})}} \tag{9}$$

This is known as one-parameter logistic (1PL) model, in another name Rasch model [25] which predicts the probability of a correct response from the interaction between the individual ability θ_j and the task parameter b_i. The parameter b_i is called the location parameter or the difficulty parameter. Cognitive ability experiment conducts picking a cognitive task of average difficulty (b about 0). If the subject gets it right, system might select a more difficult task. System can keep making the experiment more difficult until the student performs a task incorrectly. If the subject make mistake in the first task, system gives an easier task. Keep making the tasks easier until he/she gets a task correct. As soon as at least one task is correct and at least one task is incorrect the system computes a maximum likelihood estimate of the subject's standing on the trait. As soon as the system has a point estimate, it can compute a confidence interval, that is, a local standard error of measurement for the subject.

Latent response model, namely the Rasch model [25] predicts the probability of any response to a cognitive task given the true ability of the user. In general, user may have different levels of ability, and items (tasks) can differ in many respects—most importantly, some are easier, and some are more difficult. In a very simple item response setting, the subject x may only have dichotomous responses (1 = correct, or 0 = incorrect), let us consider P_{ij} as the probability of the correct response, where i refer to the task, and the index j refers to the subject.

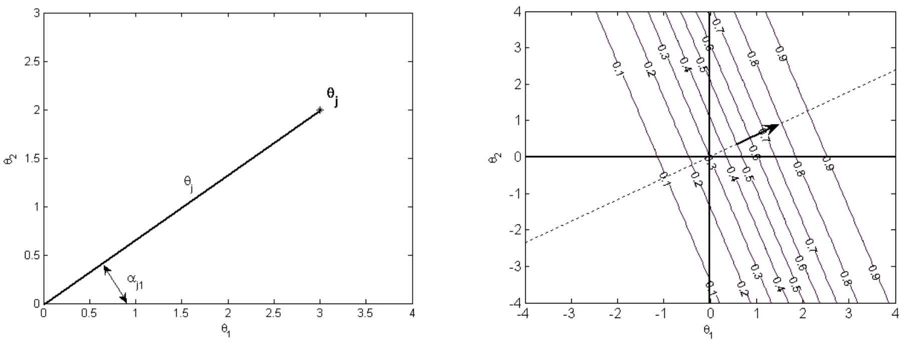
Also, $P(\theta)$ to show that the probability of a correct response is a function of the ability θ. The probability of incorrect response Q(θ) = 1 − P(θ) with Rasch modeling might show the disability, specifically the ADG [1].

$$\mathrm{Q_{ij}}\theta_j, \mathrm{b_i} = 1 - \mathrm{P_{ij}}\theta_j, \mathrm{b_i} = \frac{1}{1+\mathrm{e}^{(\theta_j - \mathrm{b_i})}} \tag{10}$$

we consider the equation (2) as gap equation. Where $\theta_j - b_i$ is considered as residue of expected ability (difficulty) and observed ability.

1) RESPONSE LATITUDE COMPUTATION

Fig. 4(B, C) illustrates category response functions [35], which are estimated to describe the likelihood that a person at a given level of the latent attitude selects a given response option. The x –axis in the Fig. 5 represents the attitude towards performing correct action (valid click or answer-



(A) (B)

FIGURE 5. **Vector representation of subject or item - (A) Polar coordinate representation of subject or item location θj, (B) Item vector representing the direction of best measurement of an item. ($a_1 = 1.2$, $a_2 = .4$, $d = -.84$).**

ing correctly), which is represented by value ranging from −4 to +4. The y-axis represents the probability that subjects at various locations along the attitude range selecting a given response option. Each response option is represented by logistic curve function running along the attitude range. The higher values along these functions indicate a higher probability of respondents selecting that particular response option.

Let, there are four b parameters associated with five point response scale. The lowest is $b_1 = -8.5$ and the highest $b_4 = 1.15$ represent the locations at which there is 50 percent probability of respondents selecting the lowest (strongly disagree) and highest (strongly agree) response options. The average b value, (.85 + .25 + .5 + 1.25)/4 = 0.7125 is considered as response latitude of the test [39]. The middle b parameters, b2 and b3, represent the intersection of middle response options. The distance between bs, shown in bracketed regions of the attitude range in the bottom of the Fig. 5B. The low distance between choices may indicate the subjects are in low load and are selective in their choice of response option.

Another important characteristic of a test item is how well it differentiates between two subjects or item located at different points in the θ-space. If the probability of the correct response to the item for the locations of two subjects is the same, the item provides no information about whether the subjects are at the same point or different points. However, if the difference in probability of the correct response is large, then it is very likely that the subjects are located at different points in the d-space. Differences in the probability of correct response for an item are largest where the slope of the item response surface is greatest, and when points in the space differ in a way that is perpendicular to the equi-probable contours for the item response surface. In this two dimensional case, $\theta_{j2} = 90 - \theta_{j1}$. More generally, the relationship among the angles between the coordinate axes and the line connecting the origin of the space to the θ_j-point is given by

$$\sum_{k=1}^{m} (\cos \alpha_{jk})^2 = 1. \tag{11}$$

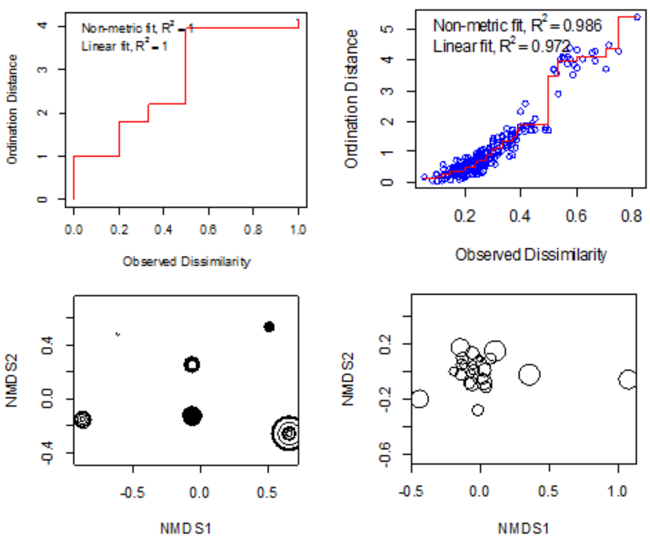


FIGURE 6. **Shepard plot of mental multiplication (top - left) and RMAP dataset (top - right). metaMDS plot of mental multiplication (bottom - left) and RMAP dataset (bottom- right).**

This relationship is a general property of the relationships of angles with a line represented in an orthogonal coordinate space. An example of this vector representation of an item is given on the contour plot of an item response surface in Fig. 6.

2) PERSON-ITEM MAP

A person-item map displays the location of item (and threshold) parameters as well as the distribution of person parameters along the latent dimension. Person-item maps are useful to compare the range and position of the item measure distribution (lower panel) to the range and position of the person measure distribution (upper panel). Items should ideally be located along the whole scale to meaningfully measure 'ability' of all persons. Fig. 12 shows, person-item map in terms of response latitude. The upper panel describes the distribution of persons' abilities and the lower panel explains item measure distributions. The black circle in the lower panel indicates mean difficulty and the white circles represent category thresholds. The left map illustrates person-item map and right map shows the sorted version.

III. RESEARCH METHOD

Item Characteristic Curve (ICC) is defined as the (nonlinear) regression line that represents the probability of endorsing an item (or an item response category) as a function of the underlying trait [29]–[41]. Though a complete description of the Item Response Theory is beyond the scope of this paper, a comprehensive analysis can be found in [33] and [47]. For the purpose of this work, we have taken into consideration the Rasch one parameter item response model (dichotomous) and Rasch extended model (polytomous data). ADG is analyzed with both dichotomous and polytomous datasets. The dichotomous data sample is considered from the pupillary dataset [45], and the polytomous data is considered from the one of our earlier experiment [1], [53]. Having both types of datasets, we performed the preprocessing, Rasch analysis and model fitting statistical tests.

A. MENTAL MULTIPLICATION DATA

There are two parts in mental multiplication dataset [26], visual and auditory. In this study, we considered only the visual task interaction part. Subjects are given three different types of tasks (easy, medium and hard). For example, 6×12 is considered as easy task, 7×19 as a medium task, and 12×17 as a hard task. The dataset is well known and followed, and prerequisites of mental multiplication task based experiments. Details can be found in [26]. The rational to use the dataset is to understand the effect of pupil size variation, eye gaze, and response time with pupil dilation. The log file gendered by the Tobii software includes different scrap values. For instance, the value marked with "-1" in pupil size means that the person is either looking away or typing or is not at the computer, meaning as the tracker is not able to detect pupil size. Some of these values are interpolated from other values preceding the current value and following values. More specifically, the interpolated pupil size is calculated as $y = x_0 + \frac{(x_1 - x_0)(t_1 - t_0)}{(t_c - t_0)}$, where tc is the time for the corresponding pupil diameter recording, x0 is preceding value of expected pupil diameter y and x1 is the following value of the expected pupil diameter. Data validation is considered from Tobii's validation values (0-4). Where, 0 represents the eye is found, and the tracking quality is good. In the case of eye out of the range, validation code is logged as 4. This study considers task interaction outcome (correct/incorrect) as subject's task performance (ability) and task difficulty (easy, medium or hard) as task demand. The pupillary data is used in validating subject's cognitive states.

B. RMAP DATASET

RMAP subjective rating dataset uses the concept of the NASA Task Load Index [15] with six dimensions to assess mental workload: mental demand, physicaldemand, temporal demand, performance, effort, and frustration. Table I shows

TABLE 1. NASA-TLX of cognitive workload computation [16].

Scale	*Description*
Mental Demand	How much mental and perceptual activity was required (e.g. thinking, deciding, calculation, remembering, looking, searching, etc.)? Was the task easy or hard, simple or complex, extracting or forgiving?
Physical Demand	How much physical activity was required (e.g., pressuring during tapping interface, tapping in different locations, double-tapping, position the camera, positioning your hand, positioning the item etc.)? Was the task easy or hard, slow or fast, slack or strenuous and restful or laborious?
Temporal Demand	How much time presser did you feel due to the rate or pace at which the task or task element occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	How successful do you think you were in accomplishing the goals of the task set by experimenter? How satisfied were you with your performance in accomplishing these goals?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration level	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

the description of NASA-TLX dimensions. Five step graded response scales are used to obtain ratings for these dimensions. A score from 0 to 10 is obtained on each scale. The six individual scale rating are combined using a weighting procedure. A cumulative workload score from 0 to 1is obtained for each rated task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by individual average score we normalized the score.

C. GAP COMPUTATION

If ICCs of two populations (group of subjects) are the same, the item is not biased. If the ICCs are different, the item is biased – which is functioning differently across the group. System can understand ability and gaps from a number of iteration, which is explained later on. Initially, the system can start with an average difficulty ($b = 0$) item if a subject get it right; the system can select more difficult item. Accordingly, the system can keep selecting more difficult items until the subject gets an item wrong. If the subject gets the first item wrong, system might give her an easier item, and can keep making the next item easier until the subject gets an item right. With at least one item right and at least one item wrong system should compute a maximum likelihood estimate of subject's standing on the trait. Having the point estimate, the system will be able to compute a confidence interval, which is a local standard error of measurement for that subject. The system then chooses that item for the subject which is expected to provide the maximum information for the subject. After administrating each item, system computes a subject's standing on the trait and his confidence interval. This we term as 'gap' in interaction. When the confidence interval (gap) is small enough, the system should stop testing. This will ensure that, each subject (with the ability) is likely to get a different

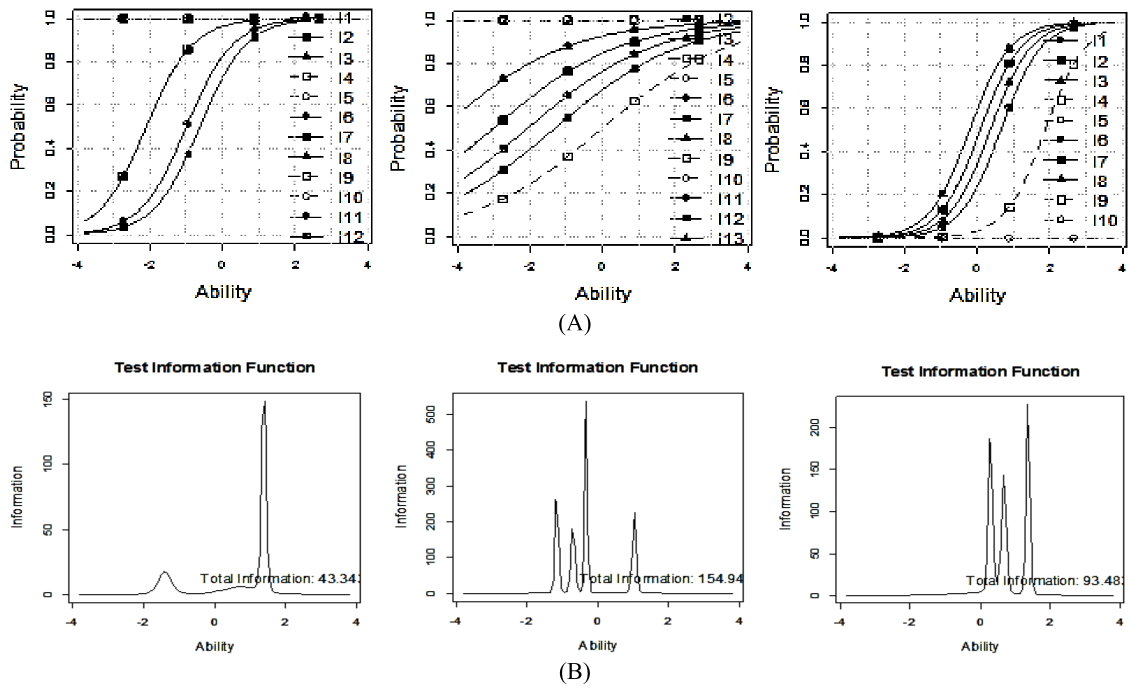


FIGURE 7. **Rasch analysis plots for mental multiplication tasks. (Left column) easy task, (middle column) medium task, and (right column) hard task - (A) Item Characteristic Curves (ICCs), (B) Test information curves.**

test but that the score will be on the same scale and measured with approximately equal error.

First, cognitive ability was computed from the residues - difference between observed proportion correct for a given ability (theta) level and expected proportion correct (i.e., probability from the IRT model). Rasch modeling and extended Rash modeling are used. Standardized residual was computed from the standard error of measurement (SEM).

Whether an item is biased or not biased is computed by the probability of an item response depends on the combination of values x of the variable X and values g of the variable G with the equation: $P(x = 1 \mid G, \theta) \neq P(x = 1|\theta)$ -which is regarded as differential item functioning (DIF). Two popular methods are broadly used in DIF computation: (a) Raju's DFIT [45] and (b) Rasch tree [44]. According to Raju [45], the gap is defined as the difference between expected scores for the focal and reference groups – and considered as the DIF index. Similarly, the cognitive gap needs to be computed from the different clusters (group) of responses and measured through DIFs. Samjima's GRM [37] was used in graded responses. Notably, the response attitude and response latitude are accounted in ADG computation. Response latitude around the attitude is found an index of gap.

Item discrimination, response biases [29], [30], and response time was added in the model for further multidimensional processing. Principal component and principal curve (non-parametric) was adopted in gap identification through response surfaces in multidimensional response modeling. The amount of stress is used to judge the goodness of fit of an MDS solution, which is computed from the sum of squared values. The absolute values of the goodness statistic depend on the definition of the stress. In all of our experimental datasets, the stress value is very low and near the perfect fit.

The adequacy of an MDS solution is visualized with Shepard diagram [27] (Fig. 6.) to show the ordinary distances and monotone or linear fit line against original dissimilarities. In addition, Shepard diagram displays two correlations like statistics on the goodness of fit in the graph. The non-metric fit is based on stress S and defined as R2=1-S2. The linear fit is squared the correlation between fitted values and ordination distances. The mental multiplication sample data set has non-metric fit (R2 = 1) and liner fit (R2 = 1). Accordingly, the RMAP dataset shows non-metric fit (R2 = 0.986) and liner fit (R2 = 0.972).

IV. RESULTS

This research aims to find a mathematical model to identify ability-demand gap (ADG) in human computer (agent) interaction. The key assumption underlying was that, human inherent ability and task complexity both are related to human task performance. Similarly, the gap might be related to human task performance, which can be identified with the same framework. Moreover, the reliability and accuracy of the identification process should be verified with fit statistics (infit/outfit).

A. DATA PREPROCESSING

As an exploratory data analysis, we performed non-metric multidimensional scaling (MDS) to infer the dimensions of the perceptual space of subjects. The raw data entering into an

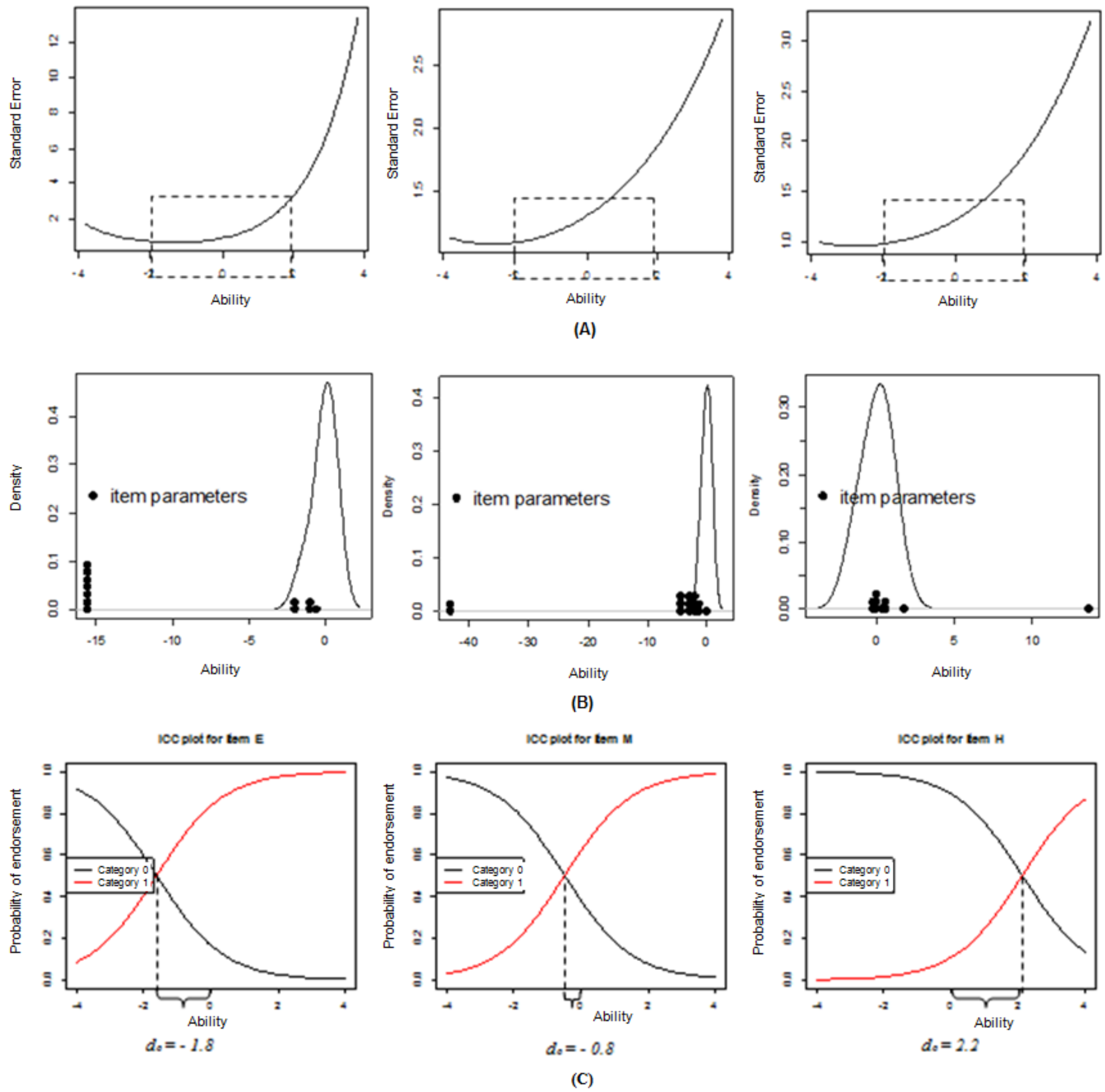


FIGURE 8. **Rasch analysis plots for mental multiplication tasks - (A) Standard Error Measurement (SEM) plot; (B) Kernel Density Estimation (KDE) plots, and (C) overall ICC plot.**

MDS analysis are typically a measure of the global similarity or dissimilarity of the stimuli or objects under investigation. A monotonic transformation of the proximities is calculated with stress function [39]. It is considered that the lesser the stress value, the better the fit of the data set [27].

Because of the nonlinear relationship between ordination and original dissimilarities, the iterative searches sometimes become very difficult in NMDS. The iteration easily gets trapped into a local optimum instead of finding the global optimum. Rotating solutions to principal components are showed from the dispersion of the points which are highest on the first dimension, using metaMDS (Fig. 6), which clearly distinguishes the dichotomous (mental multiplication data) and polytomous (RMAP dataset) responses.

B. ADG COMPUTATION FROM RESIDUAL ANALYSIS

In ADG analysis, dichotomous and polytomous datasets are processed with Rasch one parameter (1PL) item repose model and extended Rasch model, respectively.

Item characteristic, test information, item parameter, standard error of measurement and kernel density estimation of easy, medium and hard task interaction are shown in Fig. 8 and 9. Fig. 8. (A) Shows ICCs of all easy, medium, and hard (left to right) mental multiplication tasks performed by all 12 subjects. Subjects are given more medium tasks (14) then easy (12) and hard task (10). Example, of an easy task (8×12), a medium task (7×13) and a hard task (14×17). ICCs in the left part of Fig. 7A explain 12 subjects easy task interaction. Subjects correctly performed most of the easy tasks except task 1 (5×19), task 6 (7×13) and task 12

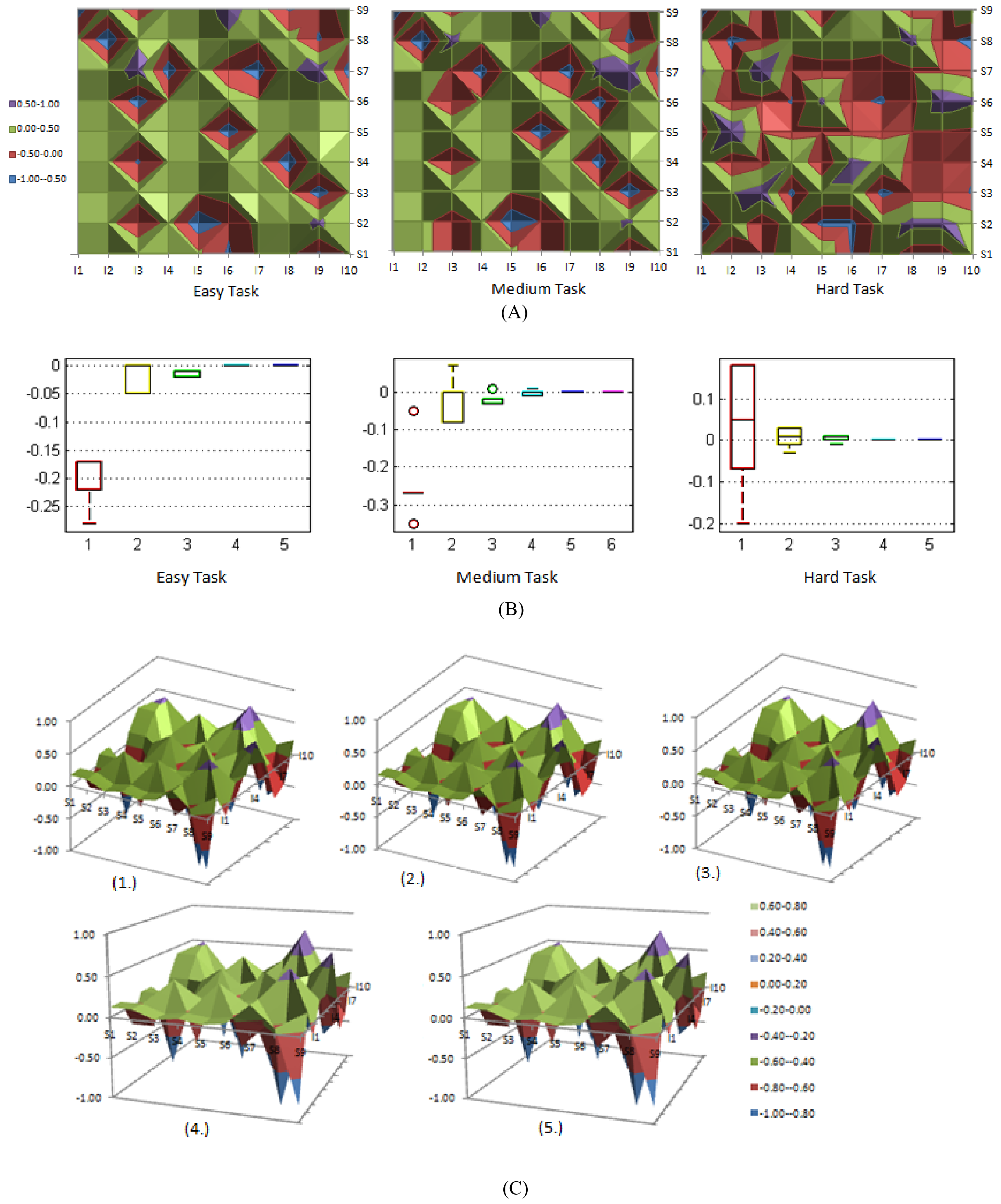


FIGURE 9. **Ability-demand gap (residuals) identification in cognitive experiment (Rasch modeling), (A) 2D surface plot of nine subjects ten tasks, (B) Rasch model convergence in easy, medium and hard task, (C) Gap convergence details in medium task.**

(9 × 17). In terms of difficulty, task 12 (19 × 13) was felt most difficult, then the task 6 (13 × 17), task 1(11 × 13) and all rest of the tasks. In terms of discrimination, easy tasks have two discrimination values 0.87 (task 1, 6, and 12). Subjects are good in guessing the outcome of other tasks, then the task 12, then task 16 and then task 1. Similarly, in ICCs of medium task (in the middle of Fig. 7B), task 8, 3, 7, 2, 4 have chronologically higher discrimination values. Except these five tasks, subjects correctly performed most of the other tasks. An overturn picture is observed in the case of

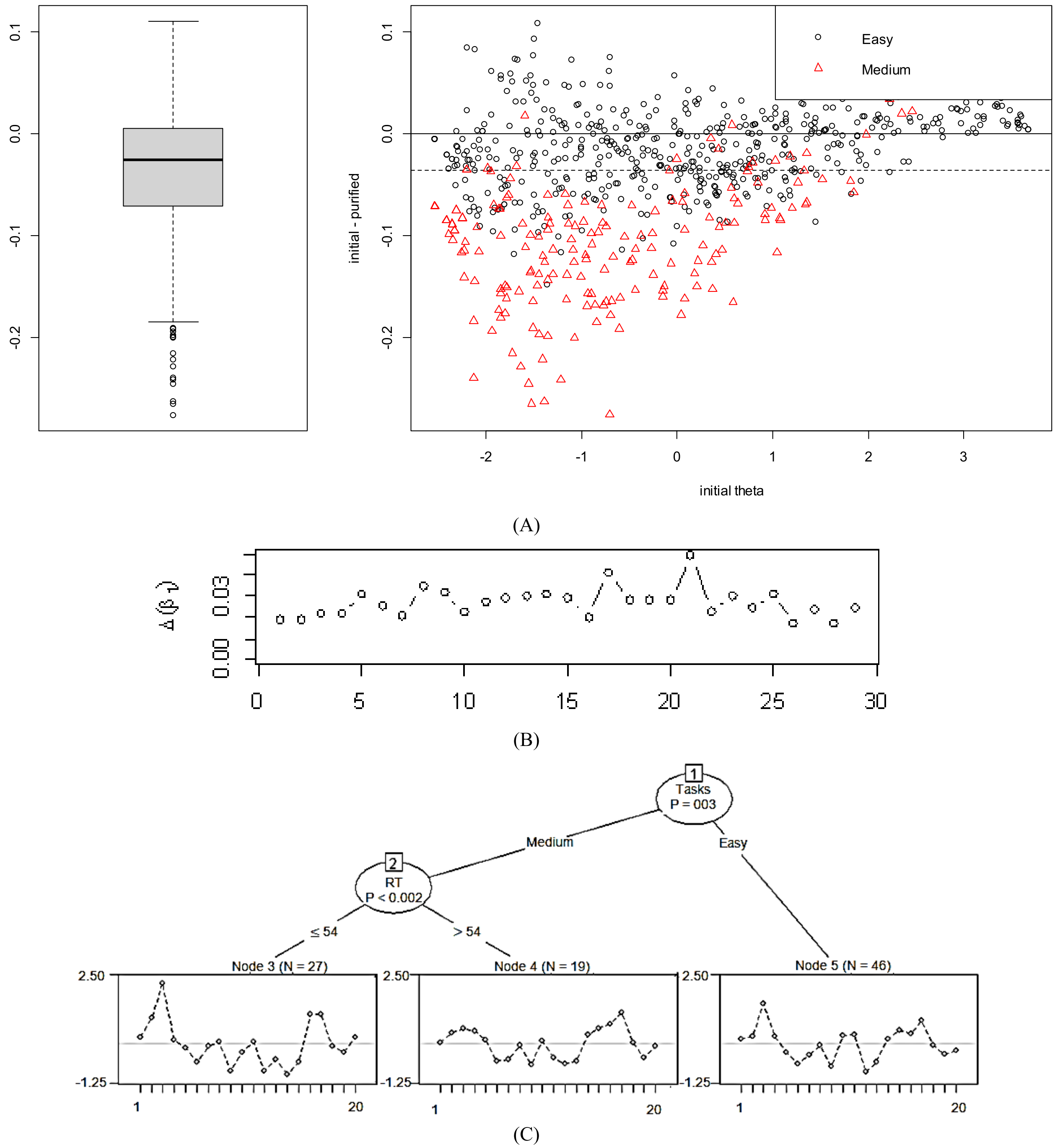


FIGURE 10. DIF plots, (A) DIF of easy vs. medium task; (B) DIF with Monte Carlo (MC) simulation, (C) DIF tree - Rasch tree plot.

hard task; almost half of the tasks are incorrectly performed by the subjects. Among five successful tasks, a few of them correctly performed the task 4. Thus, it is clear that ICCs are good enough to represent suitability of tasks in cognitive experiment. Standard error measurement (SEM) and Kernel density (KD) plots are shown in Fig. 8A and B, respectively. Although SEM plots are looks same, KD plots differentiate item density in terms of subjects' ability. The b, value in the overall ICC plot (Fig. 8C) measures the differences. The correct responses are plotted in red whereas incorrect responses are plotted using black color.

ADG is computed from average residues of endorsement in dichotomous responses. An example is shown as following. Rasch model residual analysis of easy, medium and hard tasks processed by nine subjects is shown in Fig. 9. Rasch analysis identifies users' inherent ability from a dashing point

IEEE Access
The journal for rapid open access publishing

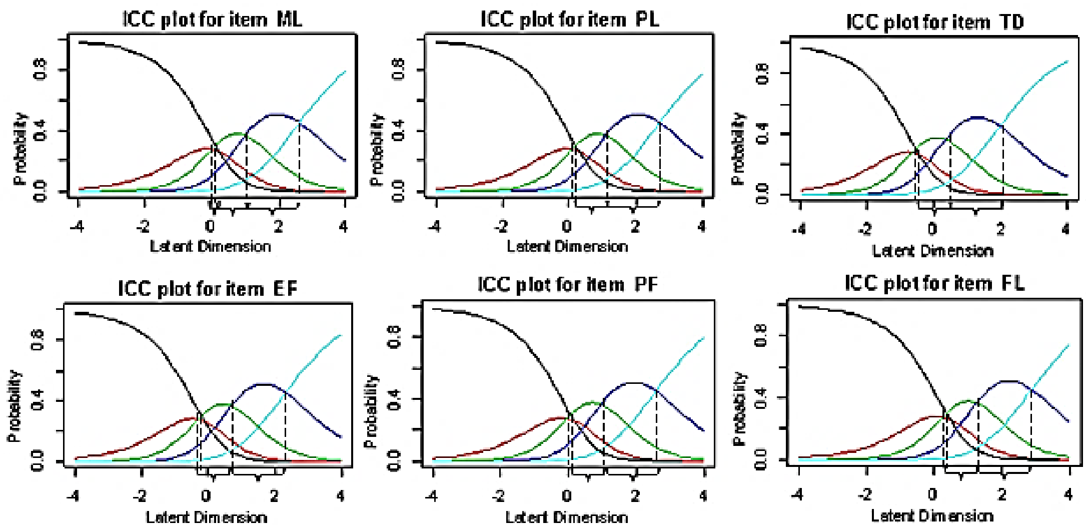


FIGURE 11. Response latitude computation from partial credit ICC plot. Top row (from left) Mental load (ML), Physical load (PL), Temporal Load (TL), Effort Factor (EF), Performance Factor (PF) and Frustration Level (FL).

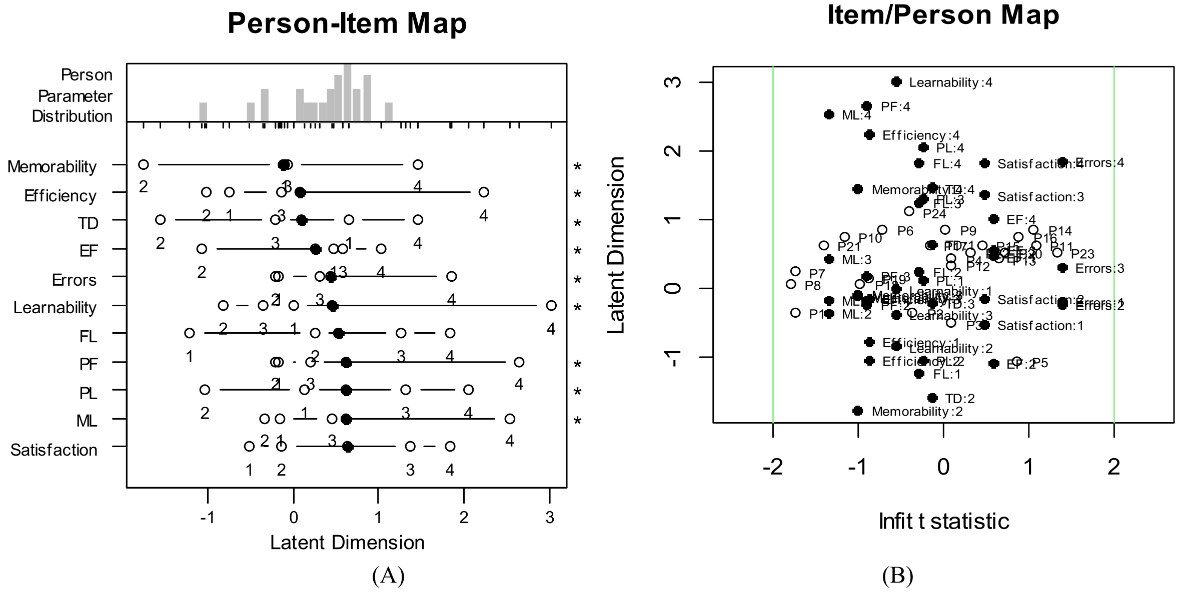


FIGURE 12. Person-Item maps- (A) Map with response latitude and (B) Map with Bond-and-Fox pathway.

of conversing the residuals (ADG) – which is considered a value very close to zero. In the example (Fig. 9B. box plots) after fifth iteration the sum of average residual converges. The hard task shows higher residuals. The medium task took relatively more iterations to converge. The X axis in the Fig. 10(b) shows the number of iteration, and the Y axis is the average ADG in terms of residuals. Fig. 9(c) illustrates the five iterations.

C. DIF ANALYSIS OF CATEGORY RELATED ADG

Rasch model necessaries that item difficulty does not change between group. For instance, subjects need more cognitive effort in medium task execution then easy tasks. It would not be surprising to see subjects failing in easy task execution response differently. Fig. 10A. Shows task level DIF impact. The box-plot (left) of the graph shows the difference in scores between using scores that ignore DIF and those that account DIF. The interquartile range, representing the middle 50% of the differences (bound between bottom and top of the shaded box), range roughly from +0.12 to −0.78 with a median of approximately +0.17. In the graph on the right, the same difference scores are plotted against the initial scores ignoring DIF (initial theta), separately for easy and medium task. Guidelines are placed at 0.0 solid line, i.e., no difference, and the mean of the differences (dotted line). The positive values to the left of the graph indicate that in almost all cases, according for DIF led, to slightly lower scores (i.e, naive score ignoring DIF minus score accounting for DIF>0, so accounting for DIF score is less than the naive score) for those with lower levels of anxiety, but this appears to be consistent across easy and medium tasks. The negative value to this graph indicates that for those with higher levels of anxiety, according for DIF led to slightly higher scores, but this again was consistent across easy and medium tasks. Higher order gap will be analyzed with 3D response modeling and robust principal component analysis. To show a prediction on large volume of data, Monte Carlo (MC) simulation is applied, which is shown in Fig. 10B. Fig. 10C, illustrates higher level classification of response tasks in terms of response time (another dimension considered with Rasch tree).

D. ADG ANALYSIS IN POLYTOMOUS RESPONSES

Ability-demand gap computation in polytomous responses is performed in RMAP dataset. The confidence interval (item-person map) and response attitude are considered as an indicator of ADG. In RMAP dataset, the ADG computed from response latitude (e.g., response latitude in mental load using absolute difference $[(|0.2–0.1| + |1.0–0.2| + |2.25–1.0|)/3 = (0.1 + 0.8 + 1.25)/3 = 0.716]$. Fig. 11 shows the item response curves of all six NASA-TLX load indexes. All load indexes shows similar response latitude ($=0.716$). The temporal load has a different orientation $[(|-(0.45)-(-0.55)| + |-(-0.55)+0.75| + |2.00-0.75|)/3 = (0.1+0.8+1.25)/3]$, but same (0.716) response latitude score. The dashed line in the Fig. 11 shows the intersection points to account response latitudes.

E. PERSON-ITEM MAP AND FIT ANALYSIS

Fit analysis is performed to ensure whether Rasch model fits the item difficulty measure or not. The 'infit' mean-square and 'outfit' mean-square statistics were performed in mental multiplication and RMAP datasets. The residual was computed by the difference between the Rasch model's theoretical explanation of item performance and the performance actually encountered for that item in the data matrix [42]. Item's dimensionality was computed by principal component analysis (PCA). It is considered that, when the variance explained by the measure is greater than 0.60 and the variance explained by the first contrast is less than 0.05, the items are unidimensional [26].

The 'infit' and 'outfit' statistics were between 0.5 to 1.5 or their standardized values between -2 to 2. The item difficulty measures analyzed by Rasch model revealed that all of the assessment items fit the model, with the 'infit' statistics ranging from 0.5 to 1.5 (Fig. 12A and Fig. 12B), but their 'outfit' t-statistics were within the range from -2 to $+2$, except for items 3, 11, 14, and 39. Only 3 items fell outside the 'infit' statistics wit criteria from 0.6 to 1.3. If the criteria were readjusted to 0.5 to 1.5, all the items fitted the 'infit' criteria. For the 'outfit' statistics, 10 of the items fell outside the range at the 0.5 to 1.5 level. However, the principal component analysis of the residual revealed that 66.1% of the variance could be explained by the model. Fig. 13B shows the Bond-and-Fox Pathway Map display of the location of each item or each person against its infit t-statistic. Pathway maps are useful for identifying misfitting items or misfitting persons. Items or people should ideally have an infit t-statistic lying between about -2 and $+2$, and these values are marked.

In the analysis of fit of mental multiplication, we found that some items did not fit the criteria of the 'infit' statistics, but we still kept them in the mental multiplication analysis for two reasons. First, all items fit the 'infit' analysis; second, they had relevance clinical meaning in the assessment. Also, we paid more attention to 'infit' and 'outfit' statistics because the later was influenced by outliers, which could easily be reminded and were less of a threat to measurement [46]. In addition, 'infit' and 'outfit' statistics adopt slightly different techniques to assess the item fit to the Rasch model. The former give more weight to the performance of persons closer to the item value whereas the later are not weighted. Therefore, the 'outfit' statistics are more sensitive to the influence of outlying score [32].

V. CONCLUSION

Understanding how cognitive ability-demand gaps (ADGs) manifest in human-machine interaction is critical in designing adaptive and assistive technology solutions. The gap or (dis) ability can be quantified with an item response model (IRT). It was observed that the IRT analysis of cognitive ability-demand gap are reliable and may be useful in shifting the paradigm of human-system (agent) interaction with application in social networks, augmented technologies, assistive technology and cyber physical systems. Also, the threshold for gap in interaction can help the system to provide relevant feedback to the user. The IRT analysis of cognitive gap can also help to design a machine that can understand group or individual information processing differences. Differential item functioning (DIF), Rash tree and principal component analysis explains more extended features of group categorical impact on gap. More research is necessary with larger datasets and variant of item and response parameters. Empirical results from this study may have broader impact in other fields of human computer interaction, assistive solutions and cognitive science.

REFERENCES

[1] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich, "Ability-based design: Concept, principles and examples," *ACM Trans. Accessible Comput.*, vol. 3, no. 3, 2011, Art. ID 9.
[2] P. Daruwalla and S. Darcy, "Personal and societal attitudes to disability," *Ann. Tourism Res.*, vol. 32, no. 3, pp. 549–570, 2005.
[3] B. Dervin, "From the mind's eye of the user: The sense-making qualitative-quantitative methodology," in *Qualitative Research in Information Management*. Englewood, CO, USA: Libraries Unlimited, 1992, pp. 61–84.
[4] J. M. Bradshaw, "Living with agents: From human-agent teamwork to cognitive prostheses," in *Intelligent Tutoring Systems*, vol. 1. Berlin, Germany: Springer-Verlag, 2002.

[5] J. M. Bradshaw, V. Dignum, C. M. Jonker, and M. Sierhuis, "Human-agent-robot teamwork," *IEEE Intell. Syst.*, vol. 27, no. 2, pp. 8–13, Mar./Apr. 2012.

[6] J. M. Bradshaw, R. R. Hoffman, D. D. Woods, and M. Johnson, "The seven deadly myths of 'autonomous systems'," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 54–61, May/Jun. 2013.

[7] K.-Y. Chung, J. Yoo, and K. J. Kim, "Recent trends on mobile computing and future networks," *Pers. Ubiquitous Comput.*, vol. 18, no. 3, pp. 489–491, 2013.

[8] M. Dunlop and S. Brewster, "The challenge of mobile devices for human computer interaction," *Pers. Ubiquitous Comput.*, vol. 6, no. 4, pp. 235–236, 2002.

[9] L. Sha, S. Gopalakrishnan, X. Liu, and Q. Wang, "Cyber-physical systems: A new frontier," in *Machine Learning in Cyber Trust*. New York, NY, USA: Springer-Verlag, 2009, pp. 3–13.

[10] Y. Tan, S. Goddard, and L. C. Pérez, "A prototype architecture for cyber-physical systems," *SIGBED Rev.*, vol. 5, no. 1, p. 26, 2008.

[11] M. C. Reddy and B. J. Jansen, "A model for understanding collaborative information behavior in context: A study of two healthcare teams," *Inf. Process. Manag.*, vol. 44, no. 1, pp. 256–273, 2008.

[12] D. A. Hofmann, Z. Lei, and A. M. Grant, "Seeking help in the shadow of doubt: The sensemaking processes underlying how nurses decide whom to ask for advice," *J. Appl. Psychol.*, vol. 94, no. 5, pp. 1261–1274, 2009.

[13] K. Alvarez and E. van Leeuwen, "To teach or to tell? Consequences of receiving help from experts and peers," *Eur. J. Social Psychol.*, vol. 41, no. 3, pp. 397–402, 2011.

[14] M. Kulkarni, "Help-seeking behaviors of people with disabilities in the workplace," *Employee Responsibil. Rights J.*, vol. 25, no. 1, pp. 41–57, 2013.

[15] C. Barrué, "Personalization and shared autonomy in assistive technologies," Ph.D. dissertation, Univ. Politècnica de Catalunya, Barcelona, Spain, 2012.

[16] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, 1988, doi: 10.1016/S0166-4115(08)62386-9.

[17] R. Azevedo *et al.*, "The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with metatutor," in *Intelligent Tutoring Systems* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2012, pp. 212–221.

[18] A. S. Shaik, G. Hossain, and M. Yeasin, "Design, development and performance evaluation of reconfigured mobile android phone for people who are blind or visually impaired," in *Proc. 28th ACM Int. Conf. Design Commun. (SIGDOC)*, 2010, pp. 159–166.

[19] N. Krahnstöver, S. Kettebekov, M. Yeasin, and R. Sharma, "iMap: A real-time framework for natural multimodal interaction with large screen displays," in *Proc. Int. Conf. Multimodal Interf.*, Pittsburgh, PA, USA, 2002, pp. 349–354.

[20] A. K. M. M. Rahman and M. Yeasin, "EmoAssist: Emotion enabled assistive tool to enhance dyadic conversation for the blind," *IEEE Trans. Affect. Comput.*, to be published.

[21] M. I. Tanveer, A. S. M. I. Anam, S. Ghosh, A. K. M. M. Rahman, and M. Yeasin, "FEPS: A sensory substitution system for the blind to perceive facial expressions," in *Proc. 14th Int. ACM SIGACCESS Conf. Comput. Access.*, Boulder, CO, USA, 2012, pp. 207–208.

[22] M. I. Tanveer, A. S. M. Anam, M. Yeasin, and M. Khan, "Do you see what I see?: Designing a sensory substitution device to access non-verbal modes of communication," in *Proc. 15th Int. ACM SIGACCESS Conf. Comput. Access.*, Oct. 2013, pp. 2–10.

[23] R. Sharma *et al.*, "Speech-gesture driven multimodal interfaces for crisis management," *Proc. IEEE*, vol. 91, no. 9, pp. 1327–1354, Sep. 2003.

[24] D. Bavelier and E. A. Hirshorn, "I see where you're hearing: How cross-modal plasticity may exploit homologous brain structures," *Nature Neurosci.*, vol. 13, pp. 1309–1311, Oct. 2010.

[25] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd ed. New York, NY, USA: Wiley, 2007.

[26] J. Klingner, B. Tversky, and P. Hanrahan, "Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks," *Psychophysiology*, vol. 48, no. 3, pp. 323–332, Mar. 2011.

[27] I. Borg and P. J. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York, NY, USA: Springer-Verlag, 2005.

[28] J. W. Thompson, "Coombs' theory of data," *Philosoph. Sci.*, vol. 33, no. 4, pp. 376–382, 1966.

[29] W. H. van Schuur, "Mokken scale analysis: Between the Guttman scale and parametric item response theory," *Political Analysis*, vol. 11, no. 2, pp. 139–163, 2003.

[30] R. J. De Ayala, *The Theory and Practice of Item Response Theory*. New York, NY, USA: Guilford Press, 2010.

[31] M. D. Reckase, "The difficulty of test items that measure more than one ability," *Appl. Psychol. Meas.*, vol. 9, no. 4, pp. 401–412, 2009.

[32] R. D. Hays, L. S. Morales, and S. P. Reise, "Item response theory and health outcomes measurement in the 21st century," *Med. Care*, vol. 38, no. 9, pp. 28–42, 2010.

[33] P. W. Holland and D. T. Thayer, "Differential item performance and the Mantel-Haenszel procedure," *Test Validity*, vol. ED 272 577, no. TM 860 493, pp. 129–145, Apr. 1988.

[34] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of Item Response Theory*. New York, NY, USA: Sage Publications, 1991.

[35] F. M. Lord, *Applications of Item Response to Theory to Practical Testing Problems*. Hillsdale, NJ, USA: Lawrence Erlbaum, 1980.

[36] S. P. Reise, and N. G. Waller, "Item response theory and clinical measurement," *Annu. Rev. Clinical Psychol.*, vol. 5, pp. 27–48, Apr. 2009.

[37] C. Strobl, J. Kopf, and A. Zeileis, "A new method for detecting differential item functioning in the Rasch model," Dept. Statist., LMU Munich, Munich, Germany, Tech. Rep. 92, 2011, vol. 1.

[38] B. D. Zumbo, *A Handbook on the Theory and Methods of Differential Item Functioning (DIF)*. Ottawa, ON, Canada: National Defense Headquarters, 1999.

[39] C. J. Lake, S. Withrow, M. J. Zickar, N. L. Wood, D. K. Dalal, and J. Bochinski, "Understanding the relation between attitude involvement and response latitude using item response theory," *Edu. Psychol. Meas.*, vol. 73, no. 4, pp. 690–712, 2013.

[40] N. S. Raju, W. J. van der Linden, and P. F. Fleer, "IRT-based internal measures of differential functioning of items and tests," *Appl. Psychol. Meas.*, vol. 19, no. 4, pp. 353–368, 1955.

[41] G. J. Mellenbergh, "Item bias and item response theory," *Int. J. Edu. Res.*, vol. 13, no. 2, pp. 127–143, 1989.

[42] R. Shealy and W. Stout, "An item response theory model for test bias," Dept. Statist., Illinois Univ. at Urbana, Urbana, IL, USA, Tech. Rep. 1991-2, 1991, vol. 2.

[43] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2011, Art. ID 11.

[44] C. Strobl, J. Kopf, and A. Zeileis, "Rasch trees: A new method for detecting differential item functioning in the Rasch model," *Psychometrika*, vol. 12, pp. 1–28, Dec. 2013.

[45] T. C. Oshima and S. B. Morris, "Raju's differential functioning of items and tests (DFIT)," *Edu. Meas., Issues Pract.*, vol. 27, no. 3, pp. 43–50, 2008.

[46] C. G. Forero and A. Maydeu-Olivares, "Estimation of IRT graded response models: Limited versus full information methods," *Psychol. Methods*, vol. 14, no. 3, pp. 275–299, 2009.

[47] S. Sinharay, "Assessing fit of unidimensional item response theory models using a Bayesian approach," *J. Edu. Meas.*, vol. 42, no. 4, pp. 375–394, 2005.

[48] P. Natesan, "A review of Bayesian item response modeling: Theory and applications," *J. Edu. Behavioral Statist.*, vol. 36, no. 4, pp. 550–552, 2007.

[49] G. Hossain and M. Yeasin, "Collaboration gaps in disabilities sensemaking: Deaf and blind communication perspective," in *Proc. ACM Conf. Comput. Supported Cooperative Work (CSCW) CIS Workshop*, San Antonio, TX, USA, 2013.

[50] G. Hossain and M. Yeasin, "Cognitive load based adaptive assistive technology design for reconfigured mobile android phone," in *Proc. 3rd Int. Conf. MobiCASE*, Los Angeles, CA, USA, 2011, pp. 370–380.

[51] G. Hossain, A. S. Shaik, and M. Yeasin, "Cognitive load and usability analysis of R-MAP for the people who are blind or visual impaired," in *Proc. 29th ACM Int. Conf. Design Commun. (SIGDOC)*, Pisa, Italy, 2011.

[52] G. Hossain and M. Yeasin, "Assistive thinking: Assistive technology design in disability management," in *Proc. 1st Int. Conf. Technol. Helping People Special Needs (ICTHP)*, Riyadh, Saudi Arabia, 2013.

[53] G. Hossain and A. H. M. A. Habib, "Designing twin 7-segment display for Bangla numeric characters," *Asian J. Inf. Technol.*, vol. 3, no. 1, pp. 49–55, 2004.

GAHANGIR HOSSAIN (S'10) is currently pursuing the Ph.D. degree in computer engineering from the Department of Electrical and Computer Engineering, University of Memphis, Memphis, TN, USA. He was an Assistant Professor with the Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh. He has research experience on human–computer interaction, cognitive vision, cyber physical systems, neurodynamics of behavior and brain, assistive technology design, and technology integration in the collaborative environment.

MOHAMMED YEASIN (SM'98) is an Associate Professor with the Department of Electrical and Computer Engineering, Adjunct Faculty Member of the Biomedical Engineering and Bioinformatics Program, and Affiliated Member of the Institute for Intelligent Systems at the University of Memphis, Memphis, TN, USA. He made significant contributions in the research and development of real-time computer vision solutions for academic research and commercial applications. He has been involved with several technological innovations, including the classifying gender, age group, ethnicity and emotion, face detection, recognition of human activities in video, and speech-gesture-enabled sophisticated natural human–computer interfaces. Some of his research on facial image analysis and hand gesture recognition is used in developing several commercial products by the Videomining Inc., State College, PA, USA.