

# Development and Validation of a Comprehensive Genomics Knowledge Scale

Michael D. Linderman<sup>a</sup> Sabrina A. Suckiel<sup>b</sup> Nathan Thompson<sup>c</sup>  
David J. Weiss<sup>c, d</sup> J. Scott Roberts<sup>e</sup> Robert C. Green<sup>f, g, h, i</sup>

<sup>a</sup>Department of Computer Science, Middlebury College, Middlebury, VT, USA; <sup>b</sup>The Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA; <sup>c</sup>Assessment Systems Corporation, Minneapolis, MN, USA; <sup>d</sup>Department of Psychology, University of Minnesota, Minneapolis, MN, USA; <sup>e</sup>Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI, USA; <sup>f</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; <sup>g</sup>Harvard Medical School, Boston, MA, USA; <sup>h</sup>The Broad Institute of MIT and Harvard, Cambridge, MA, USA; <sup>i</sup>Ariadne Labs, Boston, MA, USA

## Keywords

Genomic sequencing · Genomic literacy · Knowledge measure · Informed decision-making · Computerized adaptive testing

## Abstract

**Background:** Genomic testing is increasingly employed in clinical, research, educational, and commercial contexts. Genomic literacy is a prerequisite for the effective application of genomic testing, creating a corresponding need for validated tools to assess genomics knowledge. We sought to develop a reliable measure of genomics knowledge that incorporates modern genomic technologies and is informative for individuals with diverse backgrounds, including those with clinical/life sciences training. **Methods:** We developed the GKnowM Genomics Knowledge Scale to assess the knowledge needed to make an informed decision for genomic testing, appropriately apply genomic technologies and participate in civic decision-making. We administered the 30-item draft measure to a calibration cohort ( $n = 1,234$ ) and subsequent participants to create a combined valida-

tion cohort ( $n = 2,405$ ). We performed a multistage psychometric calibration and validation using classical test theory and item response theory (IRT) and conducted a post-hoc simulation study to evaluate the suitability of a computerized adaptive testing (CAT) implementation. **Results:** Based on exploratory factor analysis, we removed 4 of the 30 draft items. The resulting 26-item GKnowM measure has a single dominant factor. The scale internal consistency is  $\alpha = 0.85$ , and the IRT 3-PL model demonstrated good overall and item fit. Validity is demonstrated with significant correlation ( $r = 0.61$ ) with an existing genomics knowledge measure and significantly higher scores for individuals with adequate health literacy and healthcare providers (HCPs), including HCPs who work with genomic testing. The item bank is well suited to CAT, achieving high accuracy ( $r = 0.97$  with the full measure) while administering a mean of 13.5 items. **Conclusion:** GKnowM is an updated, broadly relevant, rigorously validated 26-item measure for assessing genomics knowledge that we anticipate will be useful for assessing population genomic literacy and evaluating the effectiveness of genomics educational interventions.

© 2021 S. Karger AG, Basel

## Introduction

Millions of individuals have had or will obtain personal genomic testing in educational, clinical, research, or commercial settings [1]. Achieving some degree of population genomic literacy (including provider genomic literacy) is a prerequisite to the effective application of that genomic testing [2–8]. The critical role of genomic literacy in genomic testing creates a corresponding need for validated tools to assess genomics knowledge [6].

We define genomic literacy as the knowledge needed to (1) make an informed decision for genomic testing, (2) appropriately apply genomic technologies and accurately interpret genomic data, and (3) participate in decisions about genetics and genomics policy questions as a member of society. This definition extends those described previously [9] to include the application of modern genomic technologies, the interpretation of genomic data, and the nonclinical personal, familial, and societal implications of genomic testing. The expanded definition reflects the increasing opportunities for recipients of genomic testing to obtain and analyze their own genomic data [10, 11] and the concerns of all parties about the ethical, legal, and social implications (ELSI) of genomic testing [12].

A survey of existing genetic or genomics knowledge measures is summarized in online suppl. Table 1 (see [www.karger.com/doi/10.1159/000515006](http://www.karger.com/doi/10.1159/000515006) for all online suppl. material). As genomic technology evolves, new measures are needed to address the new capabilities, limitations, and implications of those technologies. For example, the Kaphingst et al. [13] measure and the recently developed UNC Genomics Knowledge Scale (UNC-GKS) [14] and Knowledge of Genome Sequencing (KOGS) [15] measures were designed to assess genomics (vs. genetics) knowledge and to incorporate modern genomic technologies such as whole exome sequencing or whole genome sequencing (WES/WGS). Existing measures generally do not cover ELSI topics, such as the legal protections for genomic information, which are important for informed decision-making but are poorly understood [16] and/or exhibit ceiling effects in genomic study cohorts [17]. Personal genomic studies, for example, are often enriched for participants with high educational attainment and life science/health professionals in particular [18, 19]. Evaluating the effectiveness of genomics educational interventions, especially those targeting current or future healthcare providers or life scientists, requires a measure that is informative for more knowledgeable examinees.

Here we present the GKNowM Genomics Knowledge Scale, an updated and broadly relevant measure for assessing population genomic literacy and evaluating the effectiveness of genomics educational interventions. We developed GKNowM to address the need for a genomics knowledge measure that incorporates modern genomic technologies such as WES and WGS that are now in widespread use, addresses knowledge relevant to ELSI concerns, and is informative for a broad range of examinees. We performed a rigorous psychometric validation of the proposed measure with members of the general public, students/trainees, and genetics/genomics professionals, including simulations to demonstrate the suitability of the item bank for efficient computerized adaptive testing (CAT) [20]. Our goal was to create a robust and useful tool for assessing an individual's genomics knowledge across a wide range of experience, education, and expertise.

## Methods

### *Definition of Content Domain*

We developed the initial content domain based on a detailed literature review of competencies published by relevant professional societies (including the Accreditation Council for Genetic Counseling, American Board of Medical Genetics and Genomics, American College of Medical Genetics, American Nurses Association, and Association of Professors of Human and Medical Genetics), textbooks, content domains developed for existing genetics knowledge, and the investigators' professional expertise. Nine published competencies or content domains [21–28] were independently reviewed and coded for genomics concepts by 2 investigators (M.D.L. and S.A.S.); any differences were then resolved through consultation to produce the draft content domain. We revised the draft content domain based on the feedback from genomics professionals, including clinicians, researchers, and educators.

This process produced a final content domain with 3 top-level concepts: (1) "General Genetics and Genomics Knowledge," (2) "Applications of Genomics Technology," and (3) "Genomics and Society," each with multiple sub-concepts. Table 1 summarizes the content domain. For conciseness, in the remainder of the study, we will use the term "genomics" to encompass genetics and genomics.

### *Item Development*

We undertook a multistage item development process incorporating item identification, classification, adaptation and/or creation, and revision [29]. We began the item development process by reviewing 235 items from existing genetics and genomics knowledge measures, classroom exercises, and other sources (published measures are noted in online suppl. Table 1). Two investigators (M.D.L. and S.A.S.) coded the existing items with the content domain. Approximately 80% of the existing items were coded as "General Genetics and Genomics Knowledge," with fewer items coded "Applications of Genomics Technologies," and "Genomics

**Table 1.** Summary of content domain

---

General genetics and genomics knowledge
Genome composition, organization, translation, and reproduction
Genetic variation, including types, extent, constitutionality, and nomenclature
Genotype-phenotype
Inheritance
Categories, and where relevant, clinical validity, of genomic information, including health and nonhealth related
Current utility of genetic information
Application of genomic technologies
Reference genome composition and organization
Methodology, analytical capabilities, and analytical limitations of molecular testing technologies
Resources available for interpretation of genomic test results, including literature, online databases, and in silico tools
Monogenic disease variant interpretation and classification
Risk measures
Capabilities and limitations of different strategies for genomic testing, for example, trio or testing-affected individual first
Risk assessment
Genomics and society
Genomic screening
Legal protections for genetic information
Distinction between research, DTC, and clinical genomic testing
Implications and inferences for self and others of genetic information
Roles of different genomics professionals including physicians, genetic counselors, and laboratory geneticists

---

and Society.” As a result, we focused our novel item development effort in the latter 2 areas. In prior work, we observed ceiling effects with existing genomics knowledge measures [30]. To ensure that items would be informative for a broad range of examinees, including those with prior genomics education, we also focused on drafting more difficult items that would complement existing measures. The proposed measure is intended to be used in a wide variety of settings, not just in the context of a specific genomic test, such as diagnostic WES. Thus, we aimed to develop an item bank that would be relevant to multiple genomic testing technologies and contexts.

We developed an initial item bank of 74 multiple-choice items (with 3 or 4 response options) consisting of novel items and items adapted from published measures [9, 31–33]. We iteratively refined the item bank with a focus on ensuring item relevance, removing ambiguity, and clarifying the item language through feedback from genomics professionals and others. We administered the draft items to the general public recruited via Amazon Mechanical Turk and flyers, undergraduate students, healthcare professionals, and genomics professionals. Based on preliminary testing, we selected 30 high-performing items with a range of difficulties and broad coverage of the content domain. The complete text of the 30 items and the key are included in the online suppl. material.

#### *Survey*

We developed an online survey comprising demographic questions, the 30 test items, the UNC Genomics Knowledge Scale (UNC-GKS) [14], a 1-item health literacy measure [34], and a single-item self-assessment of genomics knowledge (“How would you rate your knowledge of genomics” with 5 answer options: not knowledgeable – extremely knowledgeable). We administered the

survey via the Qualtrics platform to cohorts recruited via commercial panel providers (targeting a US census panel), flyers on an undergraduate campus and at an academic medical center, an undergraduate psychology department student pool, advertisements in the National Society of Genetic Counselors e-mail bulletin, and e-mail advertisements. We recruited participants in 2 phases, an initial calibration cohort and a subsequent validation cohort. The 30 test items were also administered as part of a different study to 20 students in an intensive undergraduate human genome analysis course before and after the course (the survey was an optional element of the course). This project was reviewed and approved by the Institutional Review Boards at the Icahn School of Medicine and Middlebury College.

Respondents who completed the survey faster than a minimum time threshold (“speeders”), spent longer than 1 h, “straightlined” gridded answers, answered 85% or more of the multiple-choice items identically, or answered educational attainment demographic questions inconsistently (e.g., reported being a practicing physician without the corresponding educational attainment), were excluded from further analysis. We separately determined the minimum completion times for the calibration phase (110 s to complete the GKKnowM items) and the validation phase (330 s to complete the entire survey) based on the fastest entirely correct response to either the GKKnowM or the UNC GKS measures. We treated missing responses to GKKnowM items as incorrect.

#### *Psychometric Analysis*

Psychometric analysis was performed with a combination of R 3.5 and Xcalibre 4.2 [35]. Dimensionality and item response theory (IRT) analyses were performed on the calibration cohort, modeled on the approaches employed by Langer et al. [14] and Sanderson et al. [15].

We performed the Kaiser-Meyer-Olkin (KMO) analysis to assess suitability for factor analysis. We considered items with measure of sampling adequacy (MSA) values less than the “meritorious” threshold of 0.8 [36] for removal prior to factor analysis. We performed exploratory factor analysis (EFA) to assess the dimensionality of the items using tetrachoric correlations, a minimum residual approach, and oblimin rotation. We employed parallel analysis (PA) to identify the number of factors to be retained, generating multiple replicates ( $n = 100$ ) of randomly simulated and randomly resampled data matrices of the same size as the respondent data. Factors from the real data whose actual eigenvalues exceeded the average simulated and resampled eigenvalues were retained [37]. KMO, EFA, and PA were performed with the R psych package [38]. We performed confirmatory factor analysis (CFA) with the R lavaan package [39] using the diagonally weighted least squares estimator, robust standard errors, and mean- and variance-adjusted test statistics. We evaluated CFA model fit with root mean square error (RMSEA) (acceptable  $<0.05$ ), the standardized root mean square residual (SRMR) (acceptable  $<0.05$ ), the Tucker-Lewis index (TLI) (acceptable  $>0.95$ ), and the comparative fit index (CFI) (acceptable  $>0.95$ ) [40] and modification indices.

We performed classical test theory analyses and IRT calibration with Xcalibre. We calculated difficulty ( $p$ ) and point-biserial correlation ( $r_{pb}$ ) for each item and evaluated internal consistency with Cronbach's  $\alpha$  [41]. We fit a 3-parameter (3-PL) IRT model with discrimination ( $a$ ), difficulty ( $b$ ), and guessing ( $c$ ) parameters using a scaling constant  $D = 1.7$ . IRT scores ( $\theta$ ) were estimated with a weighted maximum likelihood approach [42] to handle all correct or all incorrect responses without assuming a prior distribution; individuals with extreme IRT scores ( $<-4$  or  $>4$ ) were excluded. We evaluated overall model fit, item fit, local dependence, and reliability with the R mirt package [43]. Overall model fit was evaluated with the  $M_2$  statistic and RMSEA, SRMR, TLI, and CFI as described above in the context of CFA. We evaluated item fit with the  $S-X^2$  goodness of fit indices [44]. To control the false discovery rate in the presence of multiple comparisons, we performed Benjamini-Hochberg (BH) correction [45, 46] when determining significance. We evaluated local dependence with the Q3 statistic of residual correlations [64].

We evaluated differential item functioning (DIF) across sex, ethnicity (white vs. non-white), and age ( $<45$  vs.  $45+$  years) to identify items that performed differently across these subgroups while holding genomics knowledge constant. For each item, we performed logistic regression with 3 models: (1) IRT score as the sole predictor, (2) IRT score and the DIF variable as predictors, and (3) IRT score, the DIF variable, and an interaction term as predictors [47]. To evaluate uniform DIF (effect is similar across the entire construct range), we performed a likelihood ratio test between models 1 and 2; to evaluate nonuniform DIF (effect differs across the construct range), we performed a likelihood ratio test between models 2 and 3. Similar to the item fit analyses, we employed the BH correction procedure when evaluating significance.

### Statistical Analysis

Statistical analyses of the GKknowM measure were performed with R 3.5 on the combined validation cohort. We evaluated concurrent validity with UNC-GKS T score [14], a previously validated genomics knowledge measure, and convergent validity with the Chew et al. [34] single question measure of health literacy. Respondents missing  $>20\%$  of the UNC-GKS items were excluded;

for the remaining respondents, missing items were scored as incorrect. We computed the Pearson correlation between the GKknowM IRT score and UNC-GKS T score and performed an independent samples unequal variances  $t$ -test to compare GKknowM IRT scores between individuals with adequate and inadequate health literacy. We predicted that the GKknowM IRT score should be positively correlated with the UNC-GKS and that individuals with adequate health literacy will have increased genomics knowledge due to their increased ability to understand health information [14].

We similarly performed an independent samples unequal variances  $t$ -test to compare GKknowM IRT scores between individuals who did and did not self-report working as or studying to be a healthcare practitioner, clinical researcher, or life scientist (abbreviated HCP) and between HCPs who did and did not report working with genomic testing/data. Since genomics is directly relevant to a HCP's professional activities, we predicted that those individuals should have increased genomics knowledge. We performed a paired  $t$ -test to evaluate the GKknowM IRT scores of students before and after they completed an intensive 1-month undergraduate course in human genome analysis (based on a similar graduate course [48]). The course material is aligned with the GKknowM construct and so we predicted that students would have significantly higher IRT scores after completing the course. We evaluated the correlation between the GKknowM IRT score and self-assessed knowledge with Kendall's Tau.

### CAT Simulation

To evaluate the utility and specification of a CAT implementation of the proposed GKknowM measure, we performed a post-hoc simulation study of the item bank [49] with the R catR package [50]. Using the existing responses of calibration cohort examinees to the fixed-form test, we “readministered” the items in simulation to identify the minimum subset of items for each examinee that resulted in IRT score estimates that were highly correlated with the IRT score from the fixed-form test.

Since the GKknowM measure is intended for measuring population genomic literacy (and not high stakes testing or other applications), we employed a static initial  $\theta$  and selected items by maximum information. The goal was to measure genomics knowledge with the minimum number of items, so we used a target standard error of measurement (SEM) as the termination criterion with a maximum number of items even if the target SEM was not reached. We simulated target SEM values of 0.30, 0.35, and 0.40, with and without a 20-item maximum limit for the calibration cohort. Given the difficulty of the items, this maximum is most likely to apply to examinees with lower knowledge levels. For this group, it is sufficient to estimate that the IRT score is much  $<0$ . We evaluated the CAT simulation with the correlation between full item bank score and CAT score estimate, the number of items, and the percentage of respondents reaching the maximum number of items.

## Results

### Sample

Of the 1,499 respondents who completed the survey during the calibration phase, 265 respondents were ex-



cluded by the quality filters. A combined total of 3,287 respondents completed the survey in the calibration and validation phases, of which 882 were excluded by the quality filters. An unknown number of respondents were excluded by the commercial cohort providers for failing to complete the survey or failing the vendor's quality filters. Table 2 lists the participant characteristics. The median time to complete the 30 draft items was 10.8 min in the calibration cohort and 10.4 min in the validation cohort after applying quality filters.

### Psychometric Analysis

Psychometric analysis and calibration were performed on the  $n = 1,234$  respondents in the calibration cohort. The overall KMO MSA was 0.92. Two items, Q22 and Q28, had KMO MSA values below the threshold of 0.80 and were removed; Q29 had a borderline MSA value (0.756) and so was retained for further analysis. The KMO results for the remaining items indicate that the data were suitable for factor analysis. In preliminary EFA with a single factor, Q26 had the smallest loading of any of the items retained after KMO analysis. Performing EFA on the subset of respondents who self-identified as HCPs, Q26 had a loading of 0.28, while all other items had loadings  $>0.40$ . Based on the factor analysis and feedback from respondents that item Q26 was confusing, we removed that item from the final measure.

PA of the remaining items indicated that 11 factors had eigenvalues greater than the eigenvalues of the random data. Online suppl. Figure 1 shows the scree plot. However, the eigenvalue of the first factor is 7.8-fold greater than the second factor indicating the presence of a single dominant factor. The actual eigenvalues for factors 3–11 are very similar to the eigenvalues for the random data; thus, we focused on a two-factor analysis. A single factor solution accounted for 30% of the variance, while the two-factor solution accounted for 34%. The second factor appeared to be associated with a single item, Q29, which had a much larger loading ( $>0.7$ ) than all other items. Based on the borderline KMO MSA, factoring, and fit issues in preliminary IRT analysis, we removed that item and proceeded with a single factor approach.

A CFA model with a single factor showed good fit ( $\chi^2 = 523.0$  [299],  $p < 0.001$ ; RMSEA = 0.025, SRMR = 0.048, CFI = 0.976, TLI = 0.974). The RMSEA and SRMR met the acceptable thresholds for good fit ( $<0.05$ ) as did CFI and TLI ( $>0.95$ ) [40]. Three pairs of items had modification indices  $>10$  suggesting local dependence (between items Q10 and Q13, Q2 and Q13, and Q17 and Q27), but

**Table 2.** Participant characteristics

	Calibration, N (%) <sup>1</sup>	Validation, N (%) <sup>1</sup>
Sex/gender		
Female	707 (57.3)	1,388 (57.7)
Male	508 (41.2)	990 (41.2)
Nonbinary	10 (0.8)	10 (0.4)
Age group		
18–24 years	251 (20.3)	361 (15.0)
25–34 years	211 (17.1)	391 (16.3)
35–44 years	193 (15.6)	413 (17.2)
45–54 years	202 (16.4)	430 (17.9)
55–64 years	175 (14.2)	396 (16.5)
65–74 years	165 (13.4)	337 (14.0)
75 or older	29 (2.1)	69 (2.9)
Race/ethnicity <sup>2</sup>		
White	869 (70.4)	1,762 (73.3)
Hispanic or Latino	182 (14.7)	290 (12.1)
Black or African American	128 (10.4)	254 (10.6)
East Asian	48 (3.9)	80 (3.3)
South Asian	23 (1.9)	53 (2.2)
Native Hawaiian or Pacific Islander	4 (0.3)	14 (0.6)
American Indian or Alaska native	15 (1.2)	45 (1.9)
Other	25 (2.0)	26 (1.1)
Highest education level completed		
Less than high school	25 (2.0)	63 (2.6)
High school graduate	287 (23.3)	549 (22.8)
Some college	322 (25.5)	624 (25.9)
Two-year college degree	123 (10.0)	260 (10.8)
Four-year college degree	204 (16.5)	438 (18.2)
Graduate work or degree	264 (21.4)	454 (18.9)
Household income		
Less than USD 40,000	452 (36.6)	911 (37.9)
USD 40,000–69,999	284 (23.0)	579 (24.1)
USD 70,000–99,999	185 (15.0)	394 (16.4)
USD 100,000–199,999	195 (15.8)	359 (14.9)
USD 200,000 or greater	102 (8.3)	134 (5.6)
Working as or studying to be a HCP		
Yes	255 (20.7)	358 (14.9)
No	974 (78.9)	2,031 (84.4)
General health literacy		
Inadequate	395 (32.0)	696 (28.9)
Adequate	814 (66.0)	1,674 (69.6)
UNC-GKS T score mean (SD)	45.5 (9.4)	45.3 (8.9)

HCP, healthcare provider; UNC-GKS, UNC Genomics Knowledge Scale. Participant characteristics for the calibration cohort ( $n = 1,234$ ) and combined validation cohort ( $n = 2,405$ ).

<sup>1</sup> Percentages may not add to 100% due to missing data.

<sup>2</sup> Percentages may sum to more than 100% because respondents may have selected more than 1 option.

**Table 3.** Item model parameters

Item	CTT		IRT 3-PL model parameters			IRT 3-PL item fit	
	<i>p</i> value	<i>r<sub>pb</sub></i>	<i>a</i>	<i>b</i>	<i>c</i>	S-X <sup>2</sup> (df)	<i>p</i> value
Q1	0.535	0.496	1.084	0.002	0.105	35.6 (16)	0.003
Q2	0.682	0.443	0.981	-0.604	0.103	11.4 (17)	0.833
Q3	0.763	0.397	1.042	-0.722	0.257	12.6 (16)	0.703
Q4	0.561	0.481	1.490	0.172	0.242	16.6 (17)	0.483
Q5	0.688	0.443	1.107	-0.362	0.247	17.9 (17)	0.395
Q6	0.628	0.426	1.041	-0.138	0.233	12.7 (17)	0.754
Q7	0.684	0.448	1.361	-0.203	0.318	11.3 (16)	0.790
Q8	0.756	0.463	1.525	-0.618	0.248	23.7 (14)	0.049
Q9	0.469	0.397	1.750	0.655	0.271	21.5 (17)	0.204
Q10	0.455	0.507	1.839	0.466	0.213	18.2 (17)	0.376
Q11	0.823	0.395	1.151	-0.942	0.325	18.1 (14)	0.202
Q12	0.443	0.429	1.168	0.630	0.217	17.3 (20)	0.635
Q13	0.402	0.292	1.637	1.047	0.274	27.7 (20)	0.118
Q14	0.525	0.415	0.979	0.320	0.230	36.6 (20)	0.013
Q15	0.592	0.414	1.002	0.065	0.243	16.2 (18)	0.575
Q16	0.334	0.228	0.998	1.569	0.244	44.3 (20)	0.001
Q17	0.594	0.322	0.721	0.136	0.257	38.5 (20)	0.008
Q18	0.444	0.425	1.660	0.662	0.245	39.4 (19)	0.004
Q19	0.386	0.399	1.294	0.898	0.220	23.7 (20)	0.258
Q20	0.371	0.197	1.088	1.537	0.280	18.1 (20)	0.581
Q21	0.414	0.372	0.881	0.860	0.212	38.2 (20)	0.008
Q23	0.468	0.292	1.751	0.963	0.350	20.2 (19)	0.382
Q24	0.459	0.461	1.726	0.532	0.234	12.6 (17)	0.762
Q25	0.316	0.316	1.657	1.294	0.227	46.2 (20)	0.001
Q27	0.413	0.298	1.271	1.080	0.274	28.7 (20)	0.093
Q30	0.422	0.388	1.089	0.775	0.222	21.4 (20)	0.372

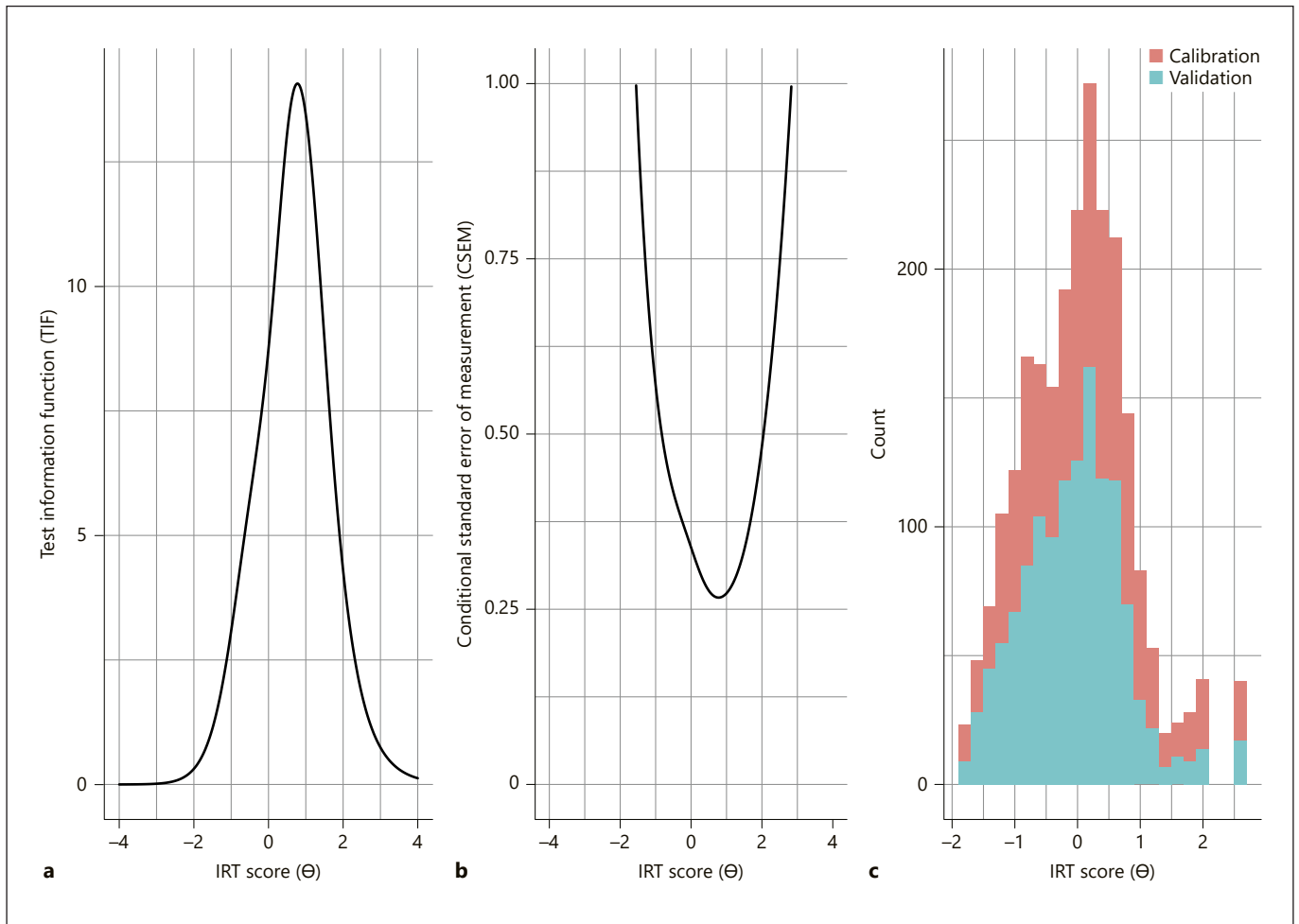
IRT 3-PL *a* parameter does not include D = 1.7 scaling constant. IRT, item response theory.

all 3 were close to the threshold (<13.3). Fitting a CFA model in which the errors for these item pairs could correlate produced a good fit ( $\chi^2 = 480.3$  [296],  $p < 0.001$ ; RMSEA = 0.022, SRMR = 0.046, CFI = 0.980, TLI = 0.978) that was statistically significantly better than the initial model ( $\chi^2 = 45.7$  [3],  $p < 0.001$ ). However, the residual correlations were <0.16, and the items had different content areas, so we retained all items.

Table 3 lists the classical test theory difficulty (*p*) and discrimination (*r<sub>pb</sub>*) results, the IRT 3-PL model parameters, and S-X<sup>2</sup> statistic for the 26-item GKnowM measure. The overall IRT 3-PL model fit met the acceptable thresholds for good fit ( $M_2 = 414.4$  [351],  $p = 0.011$ ; RMSEA = 0.012, SRMR = 0.029, TLI = 0.996, CFI = 0.996). We observed significant item misfit at the  $\alpha = 0.05$  level with BH correction for multiple items. Visual inspection of those items (Q1, Q14, Q16–Q18, Q21, Q25) indicated adequate fit, suggesting that the statistical misfit was due to sample size. Based on content relevance we

retained all items. All residual correlations were below 0.2, indicating local independence.

Online suppl. Tables 2–4 list the regression coefficients and *p* values for DIF analysis of all items. We did not observe uniform DIF for gender and ethnicity that was significant at the  $\alpha = 0.05$  level with BH correction but did observe significant nonuniform DIF for multiple items for gender and ethnicity. These results indicate that more knowledgeable white or female respondents generally performed better on these items than other individuals at the same knowledge level. However, the nonuniform DIF results may reflect demographic biases among individuals with health care, life sciences, or social sciences training. As shown in online suppl. Figure 2, many of the highest scoring respondents are, or are studying to be, genetic counselors. Genetic counselors are overwhelmingly white and female [51], and the gender and ethnicity of the high-scoring respondents in this sample reflect those demographics (with fewer nonwhite or male



**Fig. 1.** IRT analysis. **a** IRT TIF. **b** IRT CSEM function. **c** Histogram of full test IRT scores ( $\theta$ ) for both cohorts. IRT, item response theory; TIF, test information function; CSEM, conditional standard error of measurement.

individuals with higher IRT scores). When we excluded genetic counselors, we did not observe significant non-uniform DIF for gender or ethnicity. We observed uniform DIF in both directions for age for multiple items that were significant at the  $\alpha = 0.05$  level with BH correction. However, this analysis may similarly reflect the demographics of genetic counselors who are more likely to be younger than 45 years, younger undergraduate students who may be more likely to have learned about newer genomic technologies in school, and/or older individuals who may be more likely to have encountered genomic testing professionally or personally.

We performed the same DIF analysis with the combined validation cohort. We observed 1 item, Q1, to have significant uniform DIF for gender with BH correction in the larger cohort ( $\alpha < 0.001$ ). A review of the item content

did not indicate a potential explanation for gender bias and so the item was retained.

Figure 1a shows the test information function (TIF), and Figure 1b shows the conditional standard error of measurement function. The maximum TIF was 14.1 and the minimum conditional standard error of measurement was 0.267 at an IRT score of 0.78. Reliability [52] was  $>0.70$  for IRT scores of  $-1.1$  to  $2.3$ . Figure 1c shows the full test IRT score distributions for participants in the calibration and validation cohorts. The peak at higher knowledge levels reflects the HCPs, and in particular, the genomic professionals in the sample. The internal consistency of the 26-item measure was  $\alpha = 0.850$ . To facilitate applications with summed scoring, Table 4 provides a conversion between the summed score and the IRT scaled T-score [53].

**Table 4.** Conversion for summed score to IRT scaled T-score

Summed score	T-score	SE
0	31.3	5.4
1	32.0	5.4
2	32.8	5.5
3	33.7	5.6
4	34.7	5.7
5	35.8	5.7
6	37.1	5.8
7	38.5	5.7
8	40.1	5.7
9	41.8	5.5
10	43.6	5.2
11	45.4	4.8
12	47.1	4.5
13	48.7	4.1
14	50.3	3.8
15	51.7	3.5
16	53.2	3.3
17	54.6	3.1
18	55.9	3.0
19	57.3	2.9
20	58.7	2.9
21	60.2	2.9
22	61.8	3.0
23	63.6	3.2
24	65.7	3.6
25	68.4	4.1
26	72.2	5.0

Conversion computed using the calibration cohort assuming a population mean of 50 and a SD of 10. IRT, item response theory.

### Validity

The GKknowM IRT score and the previously validated UNC-GKS T score, both of which measure genomics knowledge, were significantly correlated  $r = 0.61$  (0.58–0.63),  $p < 0.001$ . Online suppl. Figure 3 shows respondents UNC-GKS T scores versus their GKknowM IRT scores. We observed a ceiling effect in which a range of higher-scoring participants answered all UNC-GKS items correctly.

Consistent with our predictions, respondents reporting adequate health literacy had higher GKknowM IRT scores ( $M = 0.10$ ) than did those reporting inadequate healthy literacy ( $M = -0.19$ ),  $t(1,439.7) = 8.02$  and  $p < 0.001$ . Similarly self-reported HCPs had higher GKknowM IRT scores ( $M = 0.53$ ) than non-HCPs ( $M = -0.08$ ),  $t(392.6) = 8.28$  and  $p < 0.001$ . Among self-reported HCPs, those individuals who reported ordering, interpreting or implementing genetic/genomic tests or using genomic

technologies in their research or employment had higher GKknowM IRT scores ( $M = 0.87$ ) than those who did not ( $M = 0.05$ ),  $t(331.7) = 6.38$  and  $p < 0.001$ .

Of the 20 students enrolled in an intensive undergraduate human genome analysis course, 14 completed the draft items before and after the course. There was a significant increase in students' GKknowM IRT scores from before ( $M = 0.93$ ) to after ( $M = 1.27$ ) the course,  $t(13) = 5.28$  and  $p < 0.001$ . To facilitate cohort comparisons when using summed scores, online suppl. Table 5 provides mean scores for the above groups using T-scores computed with the mapping in Table 4.

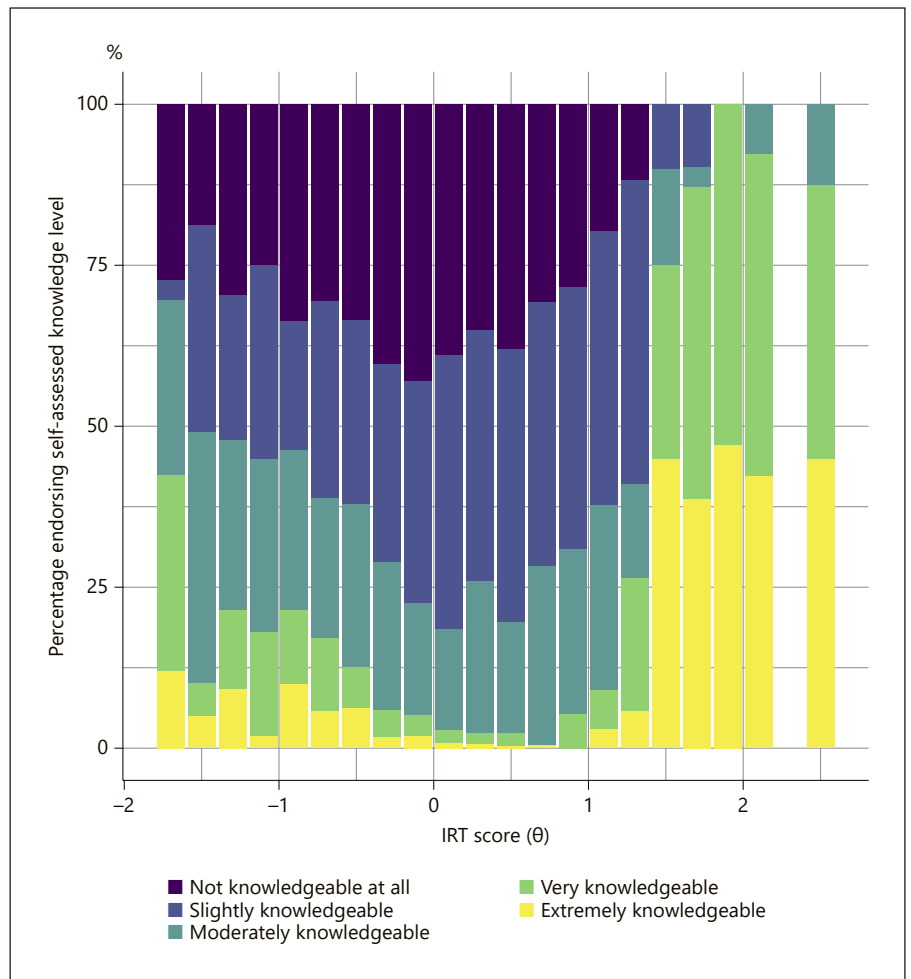
### Self-Assessed Knowledge

Figure 2 shows the percentage of individuals at each knowledge level endorsing different answers to “How would you rate your knowledge of genomics.” Individuals with higher IRT scores accurately rated themselves as “very” or “extremely” knowledgeable, while the self-assessments of individuals with lower IRT scores were less or even negatively correlated with their objectively assessed knowledge. We observed a correlation of  $\tau = 0.04$  and  $p = 0.016$  between IRT score and self-assessed knowledge for all individuals, and a significant positive correlation  $\tau = 0.29$  and  $p < 0.001$  for individuals with an IRT score  $\geq 0$  and significant negative correlation  $\tau = -0.16$  and  $p < 0.001$  for individuals with an IRT score  $< 0$ . The sharp increase in self-assessed knowledge at IRT scores  $> 1.5$  is consistent with the knowledge level above which the majority of participants are genomics professionals.

### CAT Simulation

Table 5 lists the CAT post-hoc simulation results for the calibration cohort for different target SEM values, with and without a maximum number of items. A CAT implementation with a target SEM = 0.40 and a 20-item maximum is accurate ( $r = 0.965$  compared to the full item bank) while administering approximately half the number of items on average. Incorporating a maximum number of items did not impact accuracy but reduced the test length by  $> 2$  items. Given the overall difficulty of the item bank, the maximum number of items is most likely to impact examinees at lower knowledge levels. For such a cohort, there is minimal value for administering additional items as doing so is not going to improve score accuracy.





**Fig. 2.** Self-assessed knowledge. Percentage of individuals at each full test IRT score ( $\theta$ ) endorsing different self-assessed ratings of genomics knowledge. Only bins with >10 individuals are shown. IRT, item response theory.

**Table 5.** CAT post-hoc simulation results for different target SEM values and maximum number of items

Configuration	Correlation (MSD)	Mean (SD) items	Range of items	Max. items (%)	
SEM	max.				
0.30	None	0.996 (-0.050)	23.2 (4.5)	14–26	69.2
0.30	20	0.995 (-0.050)	18.9 (2.0)	14–20	74.9
0.35	None	0.984 (-0.027)	19.2 (7.4)	8–26	51.0
0.35	20	0.983 (-0.027)	16.2 (4.7)	8–20	53.9
0.40	None	0.966 (-0.011)	15.8 (8.4)	6–26	38.2
0.40	20	0.965 (-0.011)	13.5 (5.7)	6–20	38.7

CAT, computerized adaptive testing; SEM, standard error of measurement; MSD, mean signed difference.

## Discussion

The NHGRI’s 2011 vision for the future of genomic medicine specifically cites the need for both providers and patients/consumers to achieve genomic literacy [8].

Doing so will require effective tools for assessing population genomic literacy and rigorously evaluating the effectiveness of genomics educational interventions. Existing genomics knowledge measures can be too narrowly targeted in both content and intended examinee knowledge

level. Here we present the GKNowM Genomics Knowledge Scale, a rigorously validated 26-item measure of genomic literacy. GKNowM is designed to meet the need for a genomics knowledge measure that incorporates modern genomic technologies (such as WES/WGS), is informative across a wide range of examinees (including current or future genomics professionals), and addresses the breadth of knowledge relevant to informed decision-making as both a patient and a member of society.

We anticipate the GKNowM scale will be useful in a variety of research, clinical, and educational contexts. Translational genomic studies use genomic literacy measures for multiple purposes, including to evaluate patient-focused educational interventions, such as those used in the genetic counseling protocol and as a cross-sectional measure to be tested against participant outcomes. These applications are critical and timely. The rapid advance of genomic technologies has created a corresponding need to develop new guidelines and best practices for patients/participants and providers alike. Those best practices will be informed by the data collected during the coming years on the impact of genomic literacy on the application of genomic testing and how best to promote the development of genomic literacy [6]. Similarly, genomics knowledge measures are needed to rigorously evaluate the educational interventions being developed for all stages of the educational pipeline to close the gap between the demand for and supply of genomics professionals [54–58].

These are overlapping needs. There is not always a clear distinction between the knowledge required of patients/participants versus that required of providers. For example, many consumer-facing genomic testing customers and research study participants can obtain their genomic data for further self-directed analysis and interpretation [10, 11]. Thus, in GKNowM, we sought to define a content domain that reflects the different and overlapping roles for an individual, that is, patient/participant, provider, and citizen, and an item bank that could effectively measure genomic literacy across a broad range of knowledge levels. Deploying a common measure that can be used across different settings and participant/student populations will reduce duplicated effort and facilitate comparison and meta-analysis of population genomic literacy and the educational efforts to enhance that literacy.

We performed a rigorous psychometric evaluation of GKNowM with a large, educationally, and ethnically diverse cohort drawn from the general public, students, and genomics professionals. Model fit, item fit, and dimensionality analyses indicate that GKNowM successfully measures an essentially unidimensional construct.

GKNowM items performed similarly across male/female and white/nonwhite examinees (uniform DIF), although we did observe significant nonuniform DIF for several items. The latter reflects the limited diversity in this sample at higher knowledge levels. Many of the higher-scoring participants are genetic counselors, a profession that is overwhelmingly white and female [51]. Excluding genetic counselors eliminated significant nonuniform DIF between male/female and white/nonwhite examinees. Validity analyses showed that GKNowM is significantly positively correlated with a previously validated genomics knowledge measure and a related measure of health literacy; and that groups, such as healthcare providers, whom we could expect to have increased levels of genomics knowledge, do indeed have higher scores. Thus, the GKNowM score successfully measures an individual's genomics knowledge.

GKNowM is informative for a wide range of examinees (from approximately 1 standard deviation below the mean to 2 standard deviations above the mean) and is most informative for examinees with above-average genomics knowledge. As such, it may be most useful with cohorts where participants have a range of educational backgrounds, including individuals with higher educational attainment or health or life science training [59, 60], and in evaluating educational interventions where individuals are actively learning about genomics and genome analysis. For example, we showed an application of GKNowM evaluating the knowledge gained in an undergraduate human genome analysis course (in which baseline knowledge scores were already above average). GKNowM complements existing measures, such as UNC-GKS and KOGS, which are most informative for less knowledgeable individuals (i.e., peak TIF is at IRT scores below zero). GKNowM is less informative for differentiating individuals with none versus minimal knowledge but can differentiate individuals that might otherwise “max out” other measures.

A potential limitation of the GKNowM scale is its length (26 items); the median time spent completing the original 30 online items was 10.4 min (after removing the fastest respondents). We performed simulations that showed that a CAT implementation of the measure could achieve high accuracy while administering only 13.5 items on average. A CAT approach would enjoy the benefits of the larger item bank while reducing examinee burden levels to be similar to other measures (KOGS has 9 true-false items, UNC-GKS has 19 or 25 true-false items, and Kaphingst et al. [13] have 11 items). A CAT implementation has been deployed as part of a Qualtrics survey using its survey customization tools [61] (contact the au-

thors for the customization code) and could be deployed on other online testing platforms with CAT support. Developing more time-efficient versions of GKknowM is a focus of future work. Another potential limitation is US-centric items. For example, a GKknowM item references the Genetic Information Nondiscrimination Act, a U.S. Federal Law. Users outside the USA could drop those items (noted in the online suppl. materials). As genomics technology and policy evolve some items may no longer be valid and could be dropped.

The observed relationship between self-assessed and objectively measured knowledge is consistent with previous observations that less knowledgeable individuals are less aware of that lack of knowledge [62] (what Dunning et al. [63] describe as the “double burden” of individuals’ incomplete/misguided knowledge of both the domain itself and when they are mistaken). This item was presented after the knowledge items in the survey (but before respondents could learn their score) and so would reflect their perceived performance on those items. These results may be confounded, however, by low-quality responses.

Further study is needed to evaluate the dimensionality of item subsets, multidimensional models of genomics knowledge, the stability of GKknowM scores over time, group differences at higher knowledge levels, finer grain group differences (e.g., among different racial/ethnic groups), and whether GKknowM scores are associated with differences in attitudes toward genomic testing, decision-making for genomic testing, and/or post-test psychosocial outcomes. Compared to many existing measures, the GKknowM content domain incorporates more topics focused on ELSI concerns for genomic testing. Little is known about individuals’ understanding of the scientific, legal, and other principles that underlie those concerns. Future studies of the association between genomics knowledge and ELSI concerns could inform future educational materials and counseling best practices.

## Conclusion

We present GKknowM, a rigorously validated 26-item measure for assessing genomics knowledge that is amenable to an efficient CAT implementation. The GKknowM content domain incorporates the application of modern genomic technologies, the interpretation of genomic data, and the nonclinical implications of genomic testing. This expanded definition of genomic literacy captures the knowledge needed to (1) make an informed decision for genomic testing, (2) appropriately apply genomic tech-

nologies and accurately interpret genomic data, and (3) participate in decisions about genetics and genomics policy questions as a member of society. GKknowM is an up-to-date and broadly relevant measure for assessing population genomic literacy and evaluating the pedagogical effectiveness of genomics educational interventions.

## Statement of Ethics

This project was conducted in accordance with the World Medical Association Declaration of Helsinki. Electronic written informed consent was obtained from all participants. The project was determined to be exempt by the Institutional Review Boards at the Icahn School of Medicine at Mount Sinai (#15-0355-00002-01) and Middlebury College (#16086).

## Conflict of Interest Statement

N.T. is an employee of and N.T. and D.J.W. hold an ownership interest in Assessment Systems Corporation. R.C.G. has received compensation for advising the following companies: A.I.A., Genomic Life Grail, Humanity, Kneed Media, OptumLabs, Verily, and VibrentHealth; and R.C.G. is a co-founder of Genome Medical, Inc, a technology and services company providing genetics expertise to patients, providers, employers, and care systems.

## Funding Sources

Research reported in this publication was supported by NHGRI of the National Institutes of Health under award No. R03HG008809, as well as NIH grants HD077671, HL143295, HG009922, TR003201, and HG008685, and the Franca Sozzani Fund for Preventive Genomics. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author Contributions

M.D.L. conceptualized the project and wrote the manuscript. All authors contributed to developing and refining the item bank. M.D.L. and S.A.S. performed interviews. M.D.L., S.A.S., and J.S.R. recruited survey participants. M.D.L., N.T., and D.J.W. performed the psychometric analysis. All authors have read and approved the manuscript.

## References

- 1 Birney E, Vamathevan J, Goodhand P. Genomics in healthcare: GA4GH looks to 2022. *BioRxiv*. 2017 Oct;203554.
- 2 Harvey EK, Fogel CE, Peyrot M, Christensen KD, Terry SF, McInerney JD. Providers' knowledge of genetics: a survey of 5,915 individuals and families with genetic conditions. *Genet Med*. 2007 May;9(5):259–67.
- 3 Syurina EV, Brankovic I, Probst-Hensch N, Brand A. Genome-based health literacy: a new challenge for public health genomics. *Public Health Genomics*. 2011 Jan;14(4–5):201–10.
- 4 Saul RA. Genetic and genomic literacy in pediatric primary care. *Pediatrics*. 2013 Dec;132(Suppl 3):S198–202.
- 5 Kaye C, Korf B. Genetic literacy and competency. *Pediatrics*. 2013 Dec;132(Suppl 3):S224–30.
- 6 Hurler B, Citrin T, Jenkins JF, Kaphingst KA, Lamb N, Roseman JE, et al. What does it mean to be genomically literate? National human genome research institute meeting report. *Genet Med*. 2013 Aug;15(8):658–63.
- 7 Houwink EJ, Sollie AW, Numans ME, Cornel MC. Proposed roadmap to stepwise integration of genetics in family medicine and clinical research. *Clin Transl Med*. 2013 Jan;2(1):5.
- 8 Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011 Feb;470(7333):204–13.
- 9 Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, et al. Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics*. 2008 Jan;178(1):15–22.
- 10 Ball MP, Bobe JR, Chou MF, Clegg T, Estep PW, Lunshof JE, et al. Harvard personal genome project: lessons from participatory public research. *Genome Med*. 2014 Jan;6(2):10.
- 11 Jarvik GP, Amendola LM, Berg JS, Brothers K, Clayton EW, Chung W, et al. Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *Am J Hum Genet*. 2014 May;94(6):818–26.
- 12 Robinson JO, Carroll TM, Feuerman LZ, Perry DL, Hoffman-Andrews L, Walsh RC, et al. Participants and study decliners' perspectives about the risks of participating in a clinical trial of whole genome sequencing. *J Empir Res Hum Res Ethics*. 2016 Feb;11(1):21.
- 13 Kaphingst KA, Facio FM, Cheng MR, Brooks S, Eidem H, Linn A, et al. Effects of informed consent for individual genome sequencing on relevant knowledge. *Clin Genet*. 2012 Nov;82(5):408–15.
- 14 Langer MM, Roche MI, Brewer NT, Berg JS, Khan CM, Leos C, et al. Development and validation of a genomic knowledge scale to advance informed decision making research in genomic sequencing. *MDM Policy Pract*. 2017 Jan–Jun;2(1):238146831769258.
- 15 Sanderson SC, Loe BS, Freeman M, Gabriel C, Stevenson DC, Gibbons C, et al. Development of the knowledge of genome sequencing (KOGS) questionnaire. *Patient Educ Couns*. 2018 Nov;101(11):1966–72.
- 16 Green RC, Lautenbach D, McGuire AL. GINA, genetic discrimination, and genomic medicine. *N Engl J Med*. 2015 Jan;372(5):397–9.
- 17 Suckiel SA, Linderman MD, Sanderson SC, Diaz GA, Wasserstein M, Kasarskis A, et al. Impact of genomic counseling on informed decision-making among ostensibly healthy individuals seeking personal genome sequencing: the healthseq project. *J Genet Couns*. 2016;25(5):1044–53.
- 18 Linderman MD, Nielsen DE, Green RC. Personal genome sequencing in ostensibly healthy individuals and the peopleseq consortium. *J Pers Med*. 2016;6(2):14.
- 19 Zoltick ES, Linderman MD, McGinniss MA, Ramos E, Ball MP, Church GM, et al. Predispositional genome sequencing in healthy adults: design, participant characteristics, and early outcomes of the PeopleSeq Consortium. *Genome Med*. 2019 Dec;11(1):10.
- 20 Weiss DJ. Better data from better measurements using computerized adaptive testing. *JMMSS*. 2011 Sep;2(1):1–27.
- 21 Hyland KM, Dasgupta S, Garber K, Gold J-A, Toriello H, Weissbecker K, et al. Association of professors of human and medical genetics medical school core curriculum in genetics. 2013. Available from: <http://g-2-c-2.org/resource/association-of-professors-of-human-medical-genetics>.
- 22 Accreditation Council for Genetic Counseling. Practice-Based Competencies for Genetic Counselors. 2015. Available from: [http://www.gceducation.org/Documents/ACGC\\_Core\\_Compencies\\_Brochure\\_15\\_Web.pdf](http://www.gceducation.org/Documents/ACGC_Core_Compencies_Brochure_15_Web.pdf).
- 23 Korf BR, Irons M, Watson MS. Competencies for the physician medical geneticist in the 21st century. *Genet Med*. 2011 Nov;13(11):911–2.
- 24 Hott AM, Huether CA, McInerney JD, Christianson C, Fowler R, Bender H, et al. Genetics content in introductory biology courses for non-science majors: theory and practice. *Bio-science*. 2002 Nov;52(11):1024.
- 25 American Board of Medical Genetics and Genomics. Clinical molecular genetics competencies. 2016. Available from: [http://www.ab-mgg.org/pdf/LEARNING\\_GUIDE-Molecular\\_2016.pdf](http://www.ab-mgg.org/pdf/LEARNING_GUIDE-Molecular_2016.pdf).
- 26 Marbach-Ad G. Attempting to break the code in student comprehension of genetic concepts. *J Biol Educ*. 2001;35(4):183–9.
- 27 Greco KE, Tinley S, Seibert D. Development of the essential genetic and genomic competencies for nurses with graduate degrees. *Annu Rev Nurs Res*. 2011 Jan;29:173–90.
- 28 Ward LD, Haberman M, Barbosa-Leiker C. Development and psychometric evaluation of the genomic nursing concept inventory. *J Nurs Educ*. 2014 Sep;53(9):511–8.
- 29 DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. 2007 May;45(5 Suppl 1):S12–21.
- 30 Sanderson SC, Linderman MD, Suckiel SA, Zinberg R, Wasserstein M, Kasarskis A, et al. Psychological and behavioural impact of returning personal results from whole-genome sequencing: the HealthSeq project. *Eur J Hum Genet*. 2017 Jan;25(3):280–92.
- 31 Hooker GW, Peay H, Erby L, Bayless T, Biesecker BB, Roter DL. Genetic literacy and patient perceptions of IBD testing utility and disease control: a randomized vignette study of genetic testing. *Inflamm Bowel Dis*. 2014 May;20(5):901–8.
- 32 Carere DA, Kraft P, Kaphingst KA, Roberts JS, Green RC. Consumers report lower confidence in their genetics knowledge following direct-to-consumer personal genomic testing. *Genet Med*. 2016 Jan;18(1):65–72.
- 33 Abrams LR, Koehly LM, Hooker GW, Paquin RS, Capella JN, McBride CM. Media exposure and genetic literacy skills to evaluate Angelina Jolie's decision for prophylactic mastectomy. *Public Health Genomics*. 2016;19(5):282–9.
- 34 Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008 May;23(5):561–6.
- 35 Guyer R, Thompson NA. *User's manual for Xcalibre item response theory calibration software, version 4.2.2 and later*. Woodbury, MN: Assessment Systems Corporation; 2014.
- 36 Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974 Mar;39(1):31–6.
- 37 Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ Res Methods*. 2004 Apr;7(2):191–205.
- 38 Psych RW. *Procedures for Psychological, Psychometric, and Personality Research*. 2018. Available from: <https://cran.r-project.org/package=psych>.
- 39 Rosseel Y. lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012 May;48(2):1–36.
- 40 Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. *Electron J Bus Res Methods*. 2008 Apr;6(1):53–9.
- 41 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951 Sep;16(3):297–334.
- 42 Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika*. 1989 Sep;54(3):427–50.
- 43 Chalmers RP. Mirt: a multidimensional item response theory package for the R environment. *J Stat Softw*. 2012 May;48(6):1–29.
- 44 Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. *Appl Psychol Meas*. 2000 Mar;24(1):50–64.



- 45 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;57(1):289–300.
- 46 Kim J, Oshima TC. Effect of multiple testing adjustment in differential item functioning detection. *Educ Psychol Meas*. 2013 Jun;73(3):458–70.
- 47 Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *J Educ Meas*. 1990 Dec;27(4):361–70.
- 48 Linderman MD, Bashir A, Diaz GA, Kasarskis A, Sanderson SC, Zinberg RE, et al. Preparing the next generation of genomicists: a laboratory-style course in medical genomics. *BMC Med Genomics*. 2015;8:47.
- 49 Thompson NA, Weiss DA. A framework for the development of computerized adaptive tests. *Pract Assessment Res Eval*. 2011;16(1).
- 50 Magis D, Barrada JR. Computerized adaptive testing with R : recent updates of the package catR. *J Stat Softw*. 2017 Jan;76(Code Snippet 1):1–19.
- 51 National Society of Genetic Counselors. NSGC Professional Status Survey. 2019.
- 52 Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther*. 2014 May;36(5):648–62.
- 53 Thissen D, Pommerich M, Billeaud K, Williams VSL. Item response theory for scores on tests including polytomous items with ordered responses. *Appl Psych Meas*. 1995 Mar;19(1):39–49.
- 54 McBride CM, Bowen D, Brody LC, Condit CM, Croyle RT, Gwinn M, et al. Future health applications of genomics: priorities for communication, behavioral, and social sciences research. *Am J Prev Med*. 2010 May;38(5):556–65.
- 55 Patay BA, Topol EJ. The unmet need of education in genomic medicine. *Am J Med*. 2012 Jan;125(1):5–6.
- 56 Hooker GW, Ormond KE, Sweet K, Biesecker BB. Teaching genomic counseling: preparing the genetic counseling workforce for the genomic era. *J Genet Couns*. 2014 Aug;23(4):445–51.
- 57 Korf BR, Berry AB, Limson M, Marian AJ, Murray MF, O'Rourke PP, et al. Framework for development of physician competencies in genomic medicine: report of the competencies working group of the inter-society coordinating committee for physician education in genomics. *Genet Med*. 2014 Nov;16(11):804–9.
- 58 Campion M, Goldgar C, Hopkin RJ, Prows CA, Dasgupta S. Genomic education for the next generation of health-care providers. *Genet Med*. 2019 Nov;21(11):2422–30.
- 59 Gollust SE, Gordon ES, Zayac C, Griffin G, Christman MF, Pyeritz RE, et al. Motivations and perceptions of early adopters of personalized genomics: perspectives from research participants. *Public Health Genomics*. 2012 Jan;15(1):22–30.
- 60 Green RC, Hambuch T, Ball M, Church G, Linderman M, Pearson N, et al. Predispositional genome sequencing in healthy adults: preliminary findings from the peopleseq study. *J Pers Med*. 2016 Jun;6(2):14.
- 61 Montgomery JM, Rossiter EL. So many questions, so little time: integrating adaptive inventories into public opinion research. *J Surv Stat Methodol*. 2020 Sep;8(4):667–90.
- 62 Zell E, Krizan Z. Do people have insight into their abilities? *A Metasyn PsycEXTRA Dataset*. 2014 Mar;9(2):111–25.
- 63 Dunning D. The Dunning–Kruger effect: on being ignorant of one's own ignorance. *Adv Exp Soc Psychol*. 2011 Jan;44:247–96.
- 64 Yen WM. Scaling Performance Assessments: Strategies for Managing Local Item Dependence. *J Educ Meas*. 1992 Nov;30(3):187–213.