Validation of a measure of knowledge about human papillomavirus (HPV) using item response theory and classical test theory

Jo Waller [a,*], Remo Ostini [b], Laura A.V. Marlow [a], Kirsten McCaffery [c], Gregory Zimet [d]

[a] Health Behaviour Research Centre, Department of Epidemiology and Public Health, UCL Gower Street, London WC1E 6BT, UK
[b] Healthy Communities Research Centre, Faculty of Health Sciences, The University of Queensland, Ipswich, QLD 4305, Australia
[c] Screening and Test Evaluation Program/Centre for Medical Psychology & Evidence-based Decision-making, School of Public Health, University of Sydney, NSW 2006, Australia
[d] Indiana University School of Medicine, Section of Adolescent Medicine, 410 West 10th Street, HS 1001, Indianapolis, IN 46202, USA

ARTICLE INFO

ABSTRACT

Objective. Public understanding of HPV is important to ensure informed participation in cervical cancer prevention programmes. While many studies have measured HPV knowledge, none has developed a validated measure for use across countries. We aimed to develop and validate such a measure.

Method. Items tapping knowledge of HPV, HPV testing and HPV vaccination were developed from previous literature and with expert consultation. The 29-item measure was administered via the internet to 2409 adults in the UK, US and Australia in 2011. Classical test theory and item response theory were used to establish the measure's psychometric properties.

Results. Total scale reliability was very good ($\alpha = 0.838$), as was internal consistency for a 16-item general HPV knowledge subset ($\alpha = 0.849$). Subsets of HPV testing and vaccination items showed reasonable test–retest reliability ($r_{test–retest} = 0.62$ and 0.69) but moderate internal consistency ($\alpha = 0.52$ and 0.56). Dimensionality analyses suggested that one item was not measuring the same construct as the remainder of the questionnaire. A 2-parameter logistic item response theory (IRT) model was fitted to the remaining 28 scale items.

Conclusions. A structurally coherent set of items covering a range of important HPV knowledge was developed. Responses indicated a reliable questionnaire, which allowed the fitting of an IRT model.

© 2012 Elsevier Inc. All rights reserved.

Introduction

High-risk human papillomavirus (HPV) is the primary cause of cervical cancer (Bosch and Munoz, 2002). HPV testing and vaccination are important elements for cervical cancer prevention. HPV DNA testing is increasingly used as part of cervical screening (Albrow et al., 2012; National Health and Medical Research Council Cervical Screening Guidelines Review Group, 2005; U.S.Preventive Services Task Force, 2012) and prophylactic vaccination programmes have been established in many countries (Arbyn et al., 2012).

Scientific progress in understanding HPV infection has been very rapid, but public knowledge about HPV and HPV vaccination has generally lagged behind (Dahlstrom et al., 2012; Klug et al., 2008; Li et al., 2009; Marlow et al., 2007; Samkange-Zeeb et al., 2011; Tiro et al., 2007). Low awareness and poor knowledge of HPV are concerning given the active role that people are increasingly expected to play in making decisions about their healthcare (Woolf et al., 2005). Studies have explored what women need to know about HPV from a clinical perspective (Tristram, 2006), and what women *want* to know about

HPV (Anhang et al., 2004a; McCaffery and Irwig, 2005). In addition, there is evidence that particular information about HPV (e.g. its high prevalence and likelihood of non-progression to disease) may be important in minimising the anxiety that can be associated with a positive HPV test result (Klug et al., 2008; McCaffery et al., 2006; Waller et al., 2007).

Over the last 10 years, many studies have assessed HPV knowledge in a variety of different populations and settings, using both quantitative and qualitative methods (see Klug et al., 2008 for a systematic review). However, comparison across studies is difficult, in part due to the lack of measurement consistency. There has been a tendency for each study to develop a new measure of knowledge, with minimal psychometric evaluation.

With the widespread introduction of HPV testing and vaccination, monitoring public awareness and knowledge is important, particularly with validated measures that allow comparisons across time and between populations. Such measures will facilitate the testing of public information materials to ensure that they successfully increase HPV knowledge. To this end, we developed and validated a brief measure of HPV knowledge and tested its psychometric properties, using item response theory (IRT) to model data from the UK, the US and Australia, where HPV vaccination and testing have been introduced fairly recently.

Methods

Selection of items

Items were collected from published quantitative studies of HPV knowledge (56 studies, published 1992–2009). Evaluation of these items indicated seven thematic areas into which general HPV knowledge could be grouped: 1) health consequences of HPV; 2) HPV and cervical screening; 3) symptoms; 4) causes, risk factors and transmission; 5) prevention and treatment; 6) prevalence; and 7) testing/vaccination. We selected initial items to assess general HPV knowledge that covered the first six areas, as well as items on testing and vaccination, ensuring the inclusion of issues and pieces of information found to be important to women in previous studies (Anhang et al., 2004b; McCaffery and Irwig, 2005; McCaffery et al., 2006; Waller et al., 2007). Items were discussed in detail among the authors and assessed for accuracy and completeness by two clinical HPV experts in each of the three countries (the UK, US and Australia). These steps led to the rewording, deletion and addition of some items. The final set of items included in the survey is listed in the Supplementary material. We used a 'true/false' response format, with a 'don't know' option, coded as incorrect. We used a mixture of true and false items to minimise response bias.

Participants

Participants were recruited through an online survey research company, *Survey Sampling International* (SSI), which maintains large respondent panels in multiple countries (e.g. over 1 million respondents in the US, over 200,000 in the UK, and over 80,000 in Australia). SSI sends e-mail invitations to a random subset of panel members who meet the study's entry criteria. We commissioned SSI to recruit 400 men and 400 women between the ages of 18 and 70 living in each of the three countries (USA, Australia and the UK) and to re-contact 50 male and 50 female participants who had heard of HPV from each country to complete the survey again 2-weeks later.

Procedure

From January to March 2011 potential participants were sent an invitation email with a link to the online study. Those who clicked on the link were directed to our web-survey and asked to enter their age, sex and country of residence. We set quotas to ensure we would not get more than the commissioned number of participants in each category. The online questionnaire was programmed and hosted by the Health Behaviour Research Centre, UCL and was approved by research ethics committees at UCL, University of Sydney and Indiana University.

Measures

Participants were asked 'Before today, had you ever heard of human papillomavirus (HPV)?' Those responding 'yes' were asked to complete 16 items assessing knowledge of HPV. They were then asked i) if they had heard of HPV testing, and if so, answered 6 items about HPV testing, and ii) if they had heard of HPV vaccination, and if so, answered 7 items about the vaccination. Participants also provided demographic information. The survey was completed online so there were no missing data.

Analysis

IRT was used to assess the psychometric properties of the HPV knowledge measure at the item level and to provide scale scores for respondents. In addition, classical reliability statistics and factor analyses were used to investigate test functioning and scale structure. All analyses were conducted separately on data from each country. Given very similar values for extracted factors, and reliability estimates that differed only at the second decimal place, analyses were conducted on combined data from the three countries.

Classical reliability statistics and factor analyses

We evaluated test–retest reliability and used Cronbach's alpha to assess internal consistency reliability. A high alpha also indicates the unidimensionality of the responses, and speaks to the construct validity of the test. To further evaluate dimensionality, exploratory principal axis factor analysis (EFA) and confirmatory factor analysis (CFA) were applied to item responses from each potential subscale (general HPV knowledge, HPV testing knowledge, and HPV vaccination knowledge) and from all 29 knowledge items together.

Item response theory (IRT)

IRT is a modern test theory that provides a means of evaluating and scoring response data by simultaneously modelling item and respondent characteristics, and has measurement advantages over classical methods of measurement (Ostini and Nering, 2006). The mathematical foundation of IRT is a function that relates the probability of a specific item response (e.g. a correct response to a knowledge question) to the respondent's trait level (e.g. amount of HPV knowledge) on the trait that the item is measuring (Ostini and Nering, 2006). IRT makes the following assumptions about the data produced by tests: item response data should be of known dimensionality, usually unidimensional; data must be locally independent which is usually the case if a test is unidimensional (Hambleton and Swaminathan, 1985); and finally, the latent variable is monotonically related to item response probability. Assumptions can be evaluated graphically or using item fit statistics.

A 1-parameter (Rasch-type) dichotomous IRT model (1PL) and a 2-parameter logistic IRT model (2PL) were fit to the response data using Parscale software (Muraki and Bock, 2003). The relative appropriateness of the two models for our data was evaluated by examining item fit statistics, item parameter estimate standard errors, item response function slopes, and levels of item information. Respondent knowledge scores were then estimated for each subset of items separately. Descriptive characteristics (mean and s.d.) for the three knowledge score distributions were calculated.

Finally, two additional analyses were conducted to provide validity information on the survey's functioning. First, respondent IRT scaled scores on each subset of items were correlated as a measure of overlap in the content of the three item subsets. Correlations were also calculated between responses on each subset of items and a single item survey question asking people to indicate how much they knew about HPV.

Results

Responses were obtained from 2409 participants of whom 1473 completed the general HPV knowledge items, 742 completed the HPV testing items and 1165 completed the HPV vaccination items (see Fig. 1). Sample characteristics are provided in Table 1.

Reliability analyses

Cronbach's alpha and test–retest reliability values for all 29 HPV knowledge items and for each potential subscale are provided in Table 2. Internal consistency and test–retest reliability for the 29-items were very good. The 16 general HPV knowledge items also had good internal consistency and test–retest reliability. Test–retest reliabilities for the six HPV testing items and the seven HPV vaccination items were good and fair respectively. The internal consistency of these two subscales was lower, suggesting some heterogeneity of item content.

Dimensionality analyses

The internal consistency analyses support interpreting the 16 items measuring general HPV knowledge and the full set of 29 items as conceptually coherent constructs.

Summary results from EFA investigating the unidimensionality hypothesis are shown in Table 3 and indicate good support for treating the 16 general HPV knowledge items as a single dimension. Based on the four criteria reported in Table 3, support for the unidimensionality hypothesis is the weakest for the whole 29 HPV knowledge item set. Follow-up EFA showed that removing items that did not load strongly did little to improve unidimensionality.

CFA results are presented in Table 4 and show that, for each item set, the fit of a 1-factor model is not supported from the Chi-square test and has weak support from the CFI and NFI statistics. The RMSEA and SRMR statistics provide stronger support for the unidimensionality assumption, including for the full 29 item set in the case of SRMR.
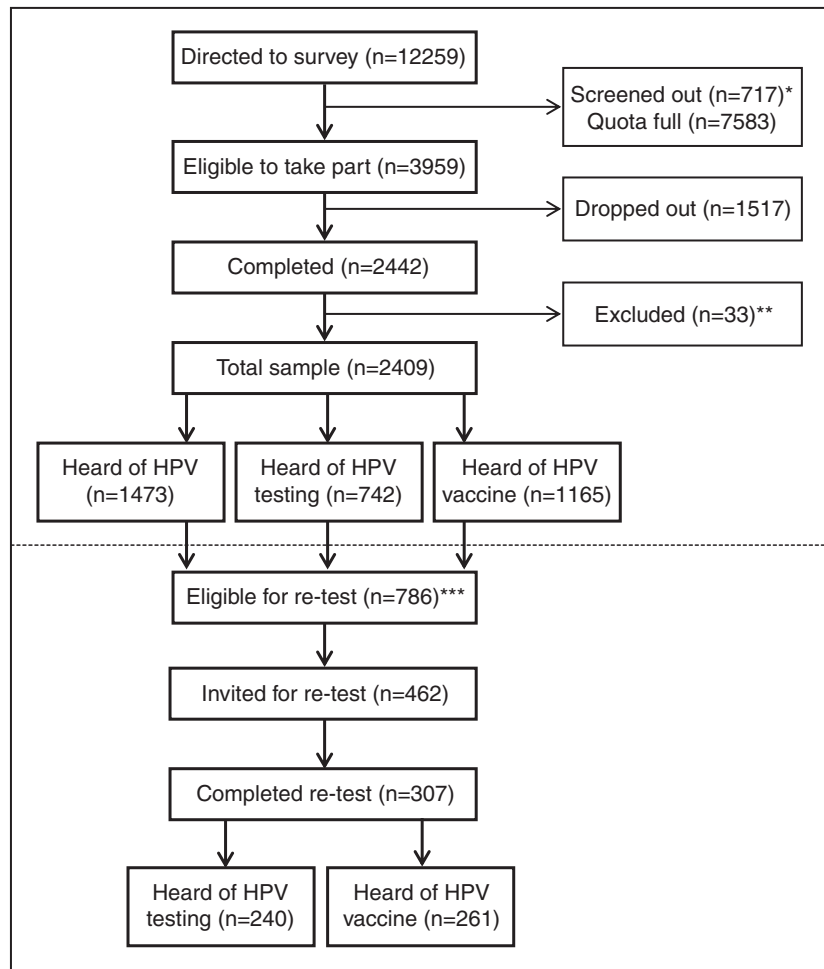
Fig. 1. Recruitment overview.

Considering the various dimensionality analyses together suggests mixed support for the unidimensionality assumption, together with weak support for factor analytic solutions with more factors. Four items (HPV causes AIDS; HPV needs no treatment; No HPV means cancer risk is low; and HPV vaccine protects against genital warts) were consistently shown to be weakly related to the remaining knowledge items. The latter three items were very difficult for respondents to answer correctly and their skewed response distributions attenuated the size of their relationships with other survey items (Bernstein and Teng, 1989). IRT's nonlinear modelling can better accommodate such skewed categorical response distributions (Bock et al., 1988). The item that asks whether HPV causes AIDS, however, appears to be measuring something different from the remaining 28 items. This combination of results suggests that a unidimensional IRT model could reasonably be fitted to a set of 28 HPV knowledge items — omitting the HPV causes AIDS item.

Table 1
Demographic characteristics of the sample included in analyses (n = 1473); HPV knowledge survey, UK, US and Australia, 2011.

	n	Age	Gender n (%)		Ethnicity n (%)		Education level n (%)		
		Mean (s.d.)	Female	Male	Majority	Minority	High	Medium	Low
United States of America	617	45.12 (15.47)	364 (51.0)	253 (49.0)	527 (85.4)	90 (14.6)	248 (40.2)	227 (36.8)	142 (23.0)
United Kingdom	404	39.96 (14.17)	249 (50.6)	155 (49.4)	332 (82.2)	72 (17.8)	178 (44.2)	149 (37.0)	76 (18.9)
Australia	452	47.87 (14.73)	288 (50.3)	164 (49.7)	347 (76.8)	105 (23.2)	120 (26.5)	165 (36.5)	167 (36.9)

Majority in US = White non-Hispanic, UK = White British, and AUS = Australian.
Education was coded as follows:
High: college graduate/graduate school (USA), degree/post-graduate degree (UK), and any university education (AUS).
Medium: some college/associate degree (USA), vocational/A-levels/other qualification<degree (UK), and vocational qualification (AUS).
Low: high school, CED or below (USA), no formal education/GCSEs (UK), and no formal education/high school (AUS).

Table 2
HPV knowledge scale reliabilities; HPV knowledge survey, UK, US and Australia, 2011.

	Alpha	Test–retest
All 29 items	0.838 (n=648)	0.79[a] (n=226)
General HPV knowledge (16 items)	0.849 (n=1473)	0.68[a] (n=307)
HPV testing knowledge (6 items)	0.521 (n=742)	0.69 (n=240)
HPV vaccination knowledge (7 items)	0.561 (n=1165)	0.62 (n=261)

[a] Retest omitted first scale item so coefficient is for 28 and 15-item scales respectively.

Item response theory analyses

Item fit test results for both Rasch and 2PL model item parameter calibrations are provided in the Supplementary material (Table S1). Setting a significance level of 0.001 to account for multiple tests and for test statistic sensitivity, this table shows better fit for the 2PL model. Item parameter estimates and associated standard errors are also provided in the Supplementary material (Table S2). The standard error results show that the Rasch parameter estimates were more precisely calculated than the 2PL model estimates (Mean Rasch location SE = 0.094; Mean 2PL location SE = 0.262).

Item response functions (for both correct and incorrect responses) are shown in Fig. 2. These plots show a wide variety of response function slopes, which together with slope parameter estimates for the 2PL model (Table S2) call into question the Rasch model assumption of uniform discrimination equal to 1.0 across all items. Finally, test information functions (TIF) for the Rasch-type model and the 2PL model (Fig. 3) show that the 2PL model provides substantially greater measurement precision across the majority of the trait scale. Greater amounts of information are provided by the 2PL in the trait range of −2.8 to 2.0 with substantially greater precision at the peak of the information function near the scaled score of −1.0.

Fig. 4 shows a histogram of the IRT scaled score distribution (mean = 0.0; s.d. = 1.0) for respondents across all 28 items. This figure shows a spike in respondents at the lowest knowledge levels. This feature of the histogram reflects the fact that a substantial number of survey respondents were unable to answer questions about

Table 4
Confirmatory factor analysis fit to 1 Factor model; HPV knowledge survey, UK, US and Australia, 2011.

	Chi square	CFI	NFI	RMSEA	SRMR
General HPV knowledge (16 items)	1981.6; p<0.0001	0.816	0.809	0.087	0.063
HPV testing knowledge (6 items)	128.0; p<0.0001	0.867	0.860	0.074	0.043
HPV vaccination knowledge (7 items)	428.9; p<0.0001	0.793	0.789	0.111	0.083
All 29 HPV knowledge items	7049.8; p<0.0001	0.620	0.601	0.086	0.081

Abbreviations: CFI: Comparative Fit Index, NFI: Normed Fit Index, RMSEA: Root Mean Square Error of Approximation, and SRMR: Standardised Root Mean Square Residual.

HPV correctly. The scaled score distribution of the general HPV knowledge subset of items had a mean of −0.234 (s.d. = 0.981), while for the HPV testing subset of items the mean was 0.000 (s.d. = 0.729) and for the HPV vaccination subset it was −0.021 (s.d. = 0.760). These results show that respondents scored best on HPV testing items and worst on general HPV knowledge.

Additional validity analyses

Intercorrelations among IRT scaled scores showed strong correlations between each subset of items. Scores on the general HPV knowledge subset of items correlated 0.502 with the HPV testing items' scores and 0.537 with respondents' scores on the vaccination items. Scores on the HPV testing and vaccination item subsets correlated 0.553. Finally, scaled scores for each subset of items also correlated significantly with a single-item measure of self-rated HPV knowledge (p<0.01 for all three item subset scores). The correlation was strongest for the general HPV knowledge item subset (r = 0.415). The relationship between self-rated knowledge about HPV and scaled scores from the HPV testing and HPV vaccination subsets of items, while still significant, was weaker (0.244, and 0.261 respectively), emphasizing the importance of objective assessment of knowledge.
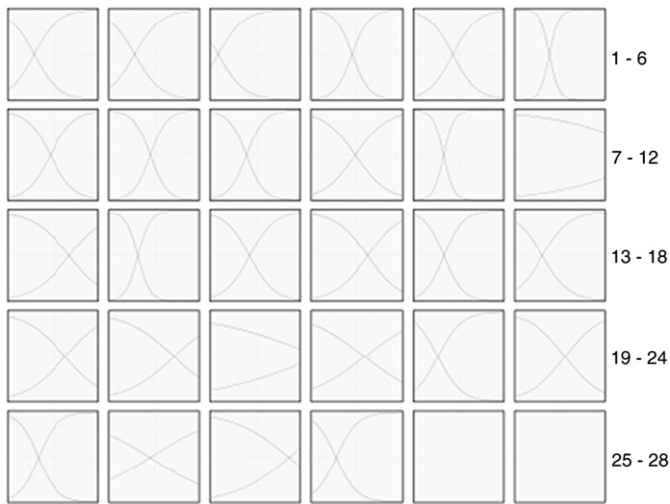
Table 3
Principal axis factor analysis (PFA) exploratory analysis of unidimensionality hypothesis; HPV knowledge survey, UK, US and Australia, 2011.

	PFA Eigen >1	F1 >3×F2	1 Factor % common variance	Load >0.33 on forced 1 factor
General HPV knowledge (16 items)	One	Yes	27.78	Two items do not: - HPV can cause HIV/AIDS - HPV usually doesn't need any treatment
HPV testing knowledge (6 items)	One	No	17.71	One item does not: - If an HPV test shows that a women does not have HPV her risk of cervical cancer is low
HPV vaccination knowledge (7 items)	One	No	21.65	Four items do not: - HPV vaccines require three doses - The HPV vaccines are most effective if given to people who have never had sex - The HPV vaccines offer protection against most cervical cancers - One of the HPV vaccines offers protection against genital warts
All HPV knowledge (29 items)	Three	Almost	20.13	Seven items do not: - HPV usually doesn't need any treatment - HPV testing is used to indicate if the HPV vaccine is needed - When you have an HPV test, you get the results the same day - If an HPV test shows that a women does not have HPV her risk of cervical cancer is low - HPV vaccines require three doses - The HPV vaccines offer protection against most cervical cancers - One of the HPV vaccines offers protection against genital warts

Fig. 2. 2-parameter logistic IRT model item response functions; HPV knowledge survey, UK, US and Australia, 2011.

Fig. 4. Histogram of HPV knowledge scaled scores across all 28 items; HPV knowledge survey, UK, US and Australia, 2011.

Discussion

This study aimed to develop a valid and reliable measure of knowledge about HPV, HPV testing and HPV vaccination suitable for use across different countries. The study benefited from inclusion of a large sample drawn from three countries, and the use of robust psychometric evaluation. The complete 29-item scale and the 16 items assessing general HPV knowledge showed good reliability on classical test theory indices. The items measuring knowledge of HPV testing and vaccination may be more informative as individual items rather than sub-scales.

Item response theory analyses indicated that although there was some heterogeneity within the construct being measured, this was diffuse rather than suggesting the existence of distinct concepts. Three items were very difficult for respondents to answer correctly and their skewed response distributions affect the dimensionality results; however they were retained because they covered important aspects of HPV knowledge. The item stating that HPV causes HIV/AIDS was excluded from the IRT analysis because of its poor performance in the scale. However, consideration should be given to retaining the question as a stand-alone item, since it addresses an important potential misunderstanding of HPV.

Overall, the results reported here show a sophisticated measure of HPV knowledge that covers a wide range of concepts with items in a format to which people 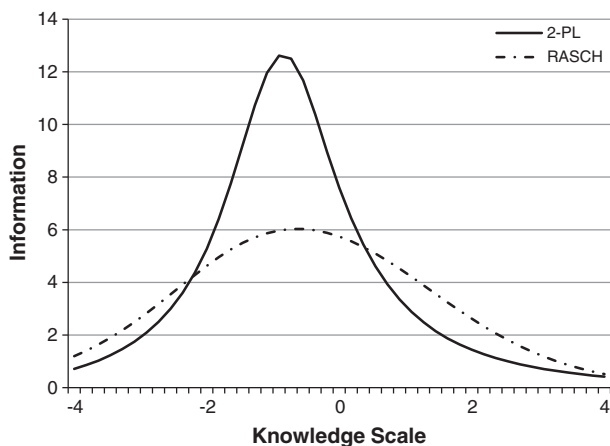are readily able to respond. The scale will be useful in collecting comparable data assessing HPV knowledge in different populations over time. This has research applications in the evaluation of public understanding of HPV but the measure could also be used by clinicians to ensure that patients undergoing HPV testing or vaccination are appropriately informed. In addition, the measure allows evaluation of the impact of public health campaigns aimed at raising public understanding of HPV. It is hoped that the publication of this validated measure will improve measurement consistency in research and practice.

The study is not without limitations. We only included three English-speaking countries and more work is needed to evaluate the validity of the measure in other settings and languages. Due to both the quota sampling method and the fact that SSI panels are not necessarily reflective of each county's demographic makeup, the representativeness of the sample is uncertain and it is likely that participants had higher than average levels of education. More validation could be needed for less-educated and low literacy groups. We anticipate that as scientific advances are made in the understanding of HPV and its relationship with cervical and other cancers, the scale may need to be updated with additional items. Extra items may also be needed to address issues that are of particular relevance to men as the vaccine becomes more available to them. However, we believe that the current version assesses aspects of HPV knowledge that are important for people making decisions about HPV testing and vaccination around the world today.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ypmed.2012.10.028.

Fig. 3. Test information functions for 1-parameter, Rasch-type IRT model and 2-parameter logistic model; HPV knowledge survey, UK, US and Australia, 2011.
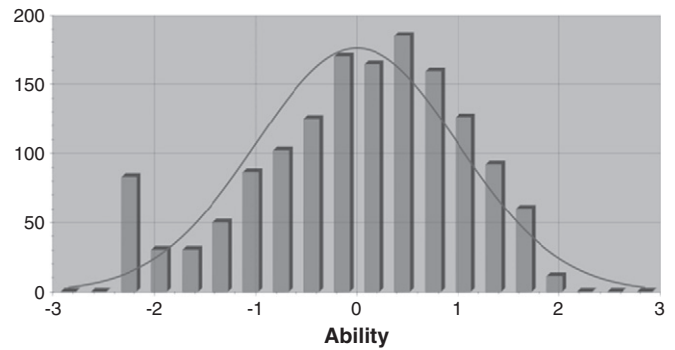
References

Albrow, R., Kitchener, H., Gupta, N., Desai, M., 2012. Cervical screening in England: the past, present, and future. Cancer Cytopathol. 120, 87–96 (available from: PM:22367883).

Anhang, R., Wright Jr., T.C., Smock, L., Goldie, S.J., 2004a. Women's desired information about human papillomavirus. Cancer 100, 315–320 (available from: PM:14716766).

Anhang, R., Wright Jr., T.C., Smock, L., Goldie, S.J., 2004b. Women's desired information about human papillomavirus. Cancer 100, 315–320.

Arbyn, M., Sanjose, S.D., Saraiya, M., et al., 2012. EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease. Int. J. Cancer 131, 1969–1982 (available from: PM:22623137).

Bernstein, I.H., Teng, G., 1989. Factoring items and factoring scales are different — spurious evidence for multidimensionality due to item categorization. Psychol. Bull. 105, 467–477 (available from: ISI:A1989U589700011).

Bock, R.D., Gibbons, R., Muraki, E., 1988. Full-information item factor-analysis. Appl. Psychol. Meas. 12, 261–280 (available from: ISI:A1988Q367900005).

Bosch, F.X., Munoz, N., 2002. The viral etiology of cervical cancer. Virus Res. 89, 183–190 (available from: PM:12445658).

Dahlstrom, L.A., Sundstrom, K., Young, C., Lundholm, C., Sparen, P., Tran, T.N., 2012. Awareness and knowledge of human papillomavirus in the Swedish adult population. J. Adolesc. Health 50, 204–206 (available from: PM:22265119).

Hambleton, R.K., Swaminathan, H., 1985. Item Response Theory: Principles and Applications. Kluwer Nijhoff, Boston.

Klug, S.J., Hukelmann, M., Blettner, M., 2008. Knowledge about infection with human papillomavirus: a systematic review. Prev. Med. 46, 87–98 (available from: PM:17942147).

Li, J., Li, L.K., Ma, J.F., et al., 2009. Knowledge and attitudes about human papillomavirus (HPV) and HPV vaccines among women living in metropolitan and rural regions of China. Vaccine 27, 1210–1215 (available from: PM:19135493).

Marlow, L.A., Waller, J., Wardle, J., 2007. Public awareness that HPV is a risk factor for cervical cancer. Br. J. Cancer 97, 691–694 (available from: PM:17687335).

McCaffery, K., Irwig, L., 2005. Australian women's needs and preferences for information about human papillomavirus in cervical screening. J. Med. Screen. 12, 134–141 (available from: PM:16156944).

McCaffery, K., Waller, J., Nazroo, J., Wardle, J., 2006. Social and psychological impact of HPV testing in cervical screening: a qualitative study. STI 82, 169–174 (available from: PM:16581749).

Muraki, E., Bock, R.D., 2003. PARSCALE 4 for Windows: IRT Based Test Scoring and Item Analysis for Graded Items and Rating Scales [Computer Software]. Scientific Software International Inc., Lincolnwood, IL.

National Health and Medical Research Council Cervical Screening Guidelines Review Group, 2005. Screening to Prevent Cervical Cancer: Guidelines for the Management of Asymptomatic Women with Screen Detected Abnormalities.

Ostini, R., Nering, M., 2006. Polytomous Item Response Theory Models. Sage, Thousand Oaks, CA.

Samkange-Zeeb, F.N., Spallek, L., Zeeb, H., 2011. Awareness and knowledge of sexually transmitted diseases (STDs) among school-going adolescents in Europe: a systematic review of published literature. BMC Publ. Health 11, 727 (available from: PM:21943100).

Tiro, J.A., Meissner, H.I., Kobrin, S., Chollette, V., 2007. What do women in the U.S. know about human papillomavirus and cervical cancer? Cancer Epidemiol. Biomarkers Prev. 16, 288–294 (available from: PM:17267388).

Tristram, A., 2006. HPV information needs. Best Pract. Res. Clin. Obstet. Gynaecol. 20, 267–277 (available from: PM:16403488).

U.S. Preventive Services Task Force, 2012. Screening for Cervical Cancer: Clinical Summary of U.S. Preventive Services Task Force Recommendation.

Waller, J., Marlow, L.A., Wardle, J., 2007. The association between knowledge of HPV and feelings of stigma, shame and anxiety. STI 83, 155–159 (available from: PM:17098767).

Woolf, S.H., Chan, E.C., Harris, R., et al., 2005. Promoting informed choice: transforming health care to dispense knowledge for decision making. Ann. Intern. Med. 143, 293–300 (available from: PM:16103473).