Development of a questionnaire for evaluating genetics education in general practice

Anna Flouris · Graeme Hawthorne ·
MaryAnne Aitken · Clara Gaff · Sylvia A. Metcalfe

Received: 6 September 2010 / Accepted: 28 September 2010 / Published online: 16 October 2010
© Springer-Verlag 2010

Abstract To support developments in genetics education, we constructed the GPGeneQ questionnaire to assess skills required for the practice of genetics by general practitioners (GPs). We describe the process of developing and validating this questionnaire to provide a detailed guide in the construction for questionnaires in the application of evaluating genetics education. The GPGeneQ was developed through a multi-step process with the initial draft based on a theoretical framework and literature review. The subsequent draft instrument contained three scales pertaining to GPs' knowledge, self-reported behaviour and attitudes regarding genetics in medicine. Content and ecological validity were measured by an iterative Delphi process involving experts, GPs and consumers of health services. Piloting to assess construct and criterion validity was conducted with a sample of GPs attending an educational workshop that was presented on a number of

separate occasions in Victoria, Australia. Results from evaluations of 145 GPs participating in ten workshops revealed evidence for validity and reliability of the GPGeneQ: knowledge change ($p < 0.001$; CI, -1.63 to -0.68), behaviour change ($p < 0.001$; CI, -4.15 to -2.21), attitudinal change ($p = 0.002$; CI, -2.68 to -0.62). This paper details the procedures involved in developing and validating an assessment questionnaire for genetics education. The GPGeneQ is the first validated questionnaire covering a broad range of topics that is designed to provide a reliable measure for the evaluation of genetics education specifically in general practice. The procedures used are transferable to the construction of any instrument for use in genetics or other medical education.

Keywords Genetics education · Questionnaire · General practice · Knowledge · Behaviour · Attitudes · Validation

Electronic supplementary material The online version of this article (doi:10.1007/s12687-010-0027-8) contains supplementary material, which is available to authorized users.

A. Flouris · M. Aitken · C. Gaff · S. A. Metcalfe (✉)
Murdoch Childrens Research Institute,
Royal Children's Hospital,
Parkville 3052 VIC, Australia
e-mail: sylvia.metcalfe@mcri.edu.au

A. Flouris · M. Aitken · C. Gaff · S. A. Metcalfe
Department of Paediatrics, The University of Melbourne,
Parkville 3052 VIC, Australia

G. Hawthorne
Department of Psychiatry, The University of Melbourne,
Parkville 3052 VIC, Australia

C. Gaff
Genetic Health Services Victoria, Royal Children's Hospital,
Parkville 3052 VIC, Australia

Introduction

Genetics is transforming the practice of clinical medicine, and family or general practitioners (GPs) are increasingly expected to play a critical frontline role in the provision of genetics services (Emery et al. 1999; Burke and Emery 2002; Sullivan et al. 2010). In many countries, including Australia, GPs generally accept their role as gatekeepers to these services, yet research to date suggests that they consider their general genetics knowledge and skills limited (Emery et al. 1999; Metcalfe et al. 2002; Suther and Goodson 2003; Baars et al. 2005). These concerns partly reflect the lag time for scientific advances to be incorporated into medical curricula and subsequently into standard medical practice where generational change is required. Consequently, continuing medical education

(CME)/continuing professional development (CPD) is especially relevant to those trained prior to the ‘genetics revolution’ since it is the premier formal mechanism for increasing GPs’ genetic literacy (Greendale and Pyeritz 2001). We acknowledge that a knowledge-deficit model is not the only aspect relevant to GPs’ practice of genetics (Kumar and Gantley 1999; Robins and Metcalfe 2004). CPD incorporates adult learning theory (ALT). ALT suggests the most effective strategies are problem-based, learner involved throughout the entire process, interactive, reflective, and multifaceted (Knowles 1984; Kaufman 2003). In Australia, CPD participation is mandatory for GPs and GP educational activities must be consistent with the principles of ALT (RACGP 2005, 2007). Following a needs assessment (Metcalf et al. 2002), we previously reported one such education program and a framework for its evaluation, also based on ALT (Gaff et al., 2007). Despite some debate in the medical education literature that certain aspects of ALT may not always apply (Newble 2002; Newman and Peile 2002), few rigorous evaluation studies have been conducted on education programs underpinned by ALT and limited instruments exist for evaluation of genetics education of GPs.

Application of the principles of evidence-based practice to the field of medical education research requires “rigorous standards of reliability and validity” (Hutchinson 1999), particularly in relation to assessment tools that are used to evaluate outcomes. While there are a number of published questionnaires used to measure the genetics-related knowledge and practice of health professionals in primary care, with a selection referenced here (Hofman et al. 1993; Kolb et al. 1999; Suther and Goodson 2003; Baars et al. 2005; Metcalfe et al. 2005; Clyman et al. 2007; Carroll et al. 2009), we are not aware of any that have been thoroughly validated and explicitly used to assess genetics education interventions in general practice that cover a broad range of topics. This study reports on the process of developing and validating an instrument for the purpose of evaluating genetics education for GPs and in particular an intervention based on ALT. This questionnaire, which we have called GPGeneQ, incorporates scales that measure knowledge, behaviour and attitudes.

Methods

Participants and ethics

This study involved three waves of participants during development and validation of the GPGeneQ, with their roles described in subsequent sections. The first wave consisted of an expert panel of professionals and representatives of consumer groups who were purposefully

recruited. Panel members were identified through their national and local profiles in these areas. The second wave consisted of a panel of practising GPs who were recruited via hospital-based GP liaison officers and academic GP departments. The third wave comprised GPs receiving the education intervention: an RACGP endorsed genetics education CPD training program. These GPs were recruited through their attendance at one of ten 2 h workshops held over a period of 3 months. Ethics approval was granted by the Human Research Ethics Committee of RACGP and all participants gave their consent prior to their inclusion in the study.

Educational intervention

The educational intervention, The Genetics File, involved two main components, a resource folder (ISBN: 0 7311 61777; www.genetichealthvic.net.au) and an interactive case-based workshop, using audio-visual materials, small and large group discussion and feedback. Briefly, the intervention was developed for GPs, and with input from GPs, based on a needs assessment and responding to greater recognition of the role of genetics in general practice (Metcalf et al. 2002). This intervention and its theoretical underpinnings have been described more fully elsewhere (Gaff et al. 2007). The workshop formed part of the CPD activities of the RACGP in the 2002–4 Quality Assurance and Professional Development (QA&PD) Triennium.

Constructing the assessment questionnaire: GPGeneQ

Construction of GPGeneQ comprised three key stages—drafting, development and validation (Table 1).

1. GPGeneQ drafting

GPGeneQ drafting involved construction of a theoretical framework describing what the assessment questionnaire aimed to measure. This was initially based on the goals of the intervention, which itself was informed by reviews of the literature, previous questionnaires developed by the authors (Metcalf et al. 2005) and consultations with GP educators.

The instrument was designed to comprise four sections including: (1) demographics and current practice; (2) genetics knowledge; (3) self-reported behaviours to assess skills; and (4) attitudes (see Electronic supplementary material ([Online resource 1](#)) for the full GPGeneQ). To account for ‘yea sayers’ both positive and negative items were included (Boynton and Greenhalgh 2004). The knowledge, behaviour and attitude sections were designed either to be scored separately or to provide a single score, representing the underlying knowledge transfer framework associated with the intervention.

Table 1 Overview of the construction of the GPGeneQ questionnaire for the drafting, development and validation stages

Process	Method	Outcome	Validity
Drafting	Literature review and consultation	Theoretical framework	N/A
	Item bank selection based on education intervention and research questions New item design and use of existing items	Item bank comprising: 19 demographic/current practice items 40 knowledge items (true/false) 9 behaviour items (open-ended) 34 attitude items	N/A
Development	Wave 1: Delphi round 1 (2 steps) ($n=11$ ‘experts’)	Modified item bank: 18 demographic/current practice items	Contribution to content and ecological validity (wave 1)
	Wave 2: Group-based questionnaire verification (GBQV) round 1 ($n=5$ GPs)	18 knowledge items (true/false/unsure) 7 behaviour items (open-ended) 27 attitude items (Likert scale)	Ecological validity (wave 2)
	Waves 1 and 2: Delphi round 2 (2 steps) ($n=7$ ‘experts’) GBQV round 2 ($n=2$ GPs)	Draft questionnaire (GPGeneQ): 20 demographic/current practice items 18 knowledge items (true/false/unsure) 7 behaviour items (open-ended) 27 attitude items (Likert scale)	Content and ecological validity (waves 1 and 2)
Validation	Wave 3: Pilot testing ($n=125$ GPs), pre-test/post-test design factor analysis/statistical analysis (see “Methods”)	GPGeneQ questionnaire: 20 demographic/current practice items 10 knowledge items (true/false/unsure), knowledge scale 7 behaviour items (open-ended): behaviour scale 16 attitude items (Likert scale): attitude scale	Construct and criterion validity (wave 3)

2. GPGeneQ development

Content validity

To establish content validity, the expert panel (wave 1) reviewed items for both content and construction using a modification of the Delphi technique, in which views of experts achieve consensus through an iterative process involving one or more rounds, each with two steps (Linstone and Turoff 1975; McLeod et al. 2004). In step 1 of the first round, a panel of experts independently reviewed the item bank and suggested modifications. The experts were also asked to:

- rate separately whether each question/item meets its objectives, and the importance and difficulty of each item on a scale from 1 (not at all: meets objectives/is important/is difficult) to 5 (extremely: meets objectives/is important/is difficult); items were removed that rated 2 (somewhat: meets objectives/is important) or

- less on the objectives and importance scales, or 4 (very difficult) or more on the difficulty scale;
- provide the most appropriate responses for the open-ended behavioural items;
- modify, omit or add items.

All possible responses were collated to provide a score for each of the items. Suggestions and scores were summarised and returned to all participants in the second step of this first round for: review and rating; ranking of the open-ended responses; and further opportunity to comment. The item bank was then modified and presented to the panel for the second Delphi round, in which this two-step process was repeated.

Ecological validity

Establishing ecological validity is essential to determine if instrument items are appropriate for use with the target population, in this case GPs (Streiner and Norman 1995). In

the current study this was addressed in three waves. Firstly, the panel experts (wave 1) included practising clinicians who were familiar with the Australian CPD process, clinical conditions and genetics education. Secondly, GPs (wave 2) were involved in initial group-based questionnaire verification (GBQV) regarding the relevance, clarity, ambiguity and difficulty of each item, scale or component, and the questionnaire as a whole. The results of the Delphi rounds and GBQV were incorporated to produce a modified item bank. The process outlined above was then repeated, resulting in yet further revisions of the questionnaire, here called the draft GPGeneQ. Thirdly, this draft GPGeneQ was pilot tested with consenting GPs (wave 3) recruited through participation in the education intervention (Gaff et al. 2007).

3. GPGeneQ validation

Edumetric and psychometric measures

Tests that primarily aim to “measure the gain or growth of individuals, for example, the measurement of knowledge, skill or achievement”, should be evaluated on edumetric principles whilst those which primarily aim to “measure individual differences, for example, a general aptitude, ability or trait” (Carver 1974, p 513) are generally evaluated using psychometric principles. The knowledge scale of draft GPGeneQ was validated as an edumetric measure, whereas the behaviour and attitude scales were validated as psychometric measures.

Sample size

Recommended sample sizes for initial tests of validity and reliability vary widely, with some authors suggesting a specific sample size (De Vaus 1995), and others suggesting that the ratio of respondents to items to undergo factor analysis is more important (Nunnally 1978; Tabachnick and Fidell 2001). In this study, ten edumetric items and 23 psychometric items were piloted using Guadagnoli’s recommendation of approximately 150 respondents for stable factor analysis (Guadagnoli and Velicer 1988).

Item selection criteria: construct and criterion validity

Range endorsement, redundancy (double-counting) and missing data were the criteria used to establish construct and criterion validity for items. Range endorsement refers to respondents selecting all possible responses for an item (other than attitude items). An exception for attitude items was made because GPs in the GBQV argued that GPs would generally feel favourable towards genetics unless they had personally had a bad experience.

For each item within the edumetric knowledge scale, pilot workshop participants were awarded a point for answering correctly, no points for an ‘unsure’ answer and a point was deducted for an incorrect answer. This stringent scoring scheme allowed differentiation between incorrect and unsure responses, and to minimise responders guessing the answer. Those items with the poorest features were removed, i.e. those which were repetitious, which failed to discriminate between known groups as expected and those which had more than 5% missing values. Responses to surviving items were then summed to produce a total score for knowledge, ranging from +10 to –10. The criterion for discrimination was a positive, statistically significant change in scores when measured before and after workshop participation.

Validation of the psychometric scales (behaviour and attitude) involved factor analysis, specifically principal components analysis (PCA), in which the criterion for retention in the scale was the conventional loading of 0.30 or greater (Rummel 1970). A second criterion was reliability assessed by internal consistency using Cronbach’s α and item-total-correlations (r_{it}). Items with an α score of at least 0.6 and r_{it} of greater than 0.2 were included in the scale (Cortina 1993).

Statistical data analysis

All analyses were performed using the statistical package SPSS for Windows, Version 11.5 (Statistical Program for the Social Sciences, SPSS Inc, Chicago, IL). Descriptive statistics such as frequencies and means were calculated for each item response, and scores were derived for each of the knowledge, behaviour and attitude scales. Results are expressed as a percentage of those who responded to each item (valid%). Differences between demographic and practice characteristics were measured using the Chi-square (χ^2) test for independent samples. Longitudinal differences between scores were measured using paired sample t tests. A p value <0.05 was considered statistically significant.

Results

Participants

All invited 11 experts (wave 1) in the field of genetics education, clinical genetics, genetic counselling, psychometrics, public health genetics, and representatives of consumer groups participated in the first Delphi round to develop the item bank. Of these, seven completed the second Delphi round contributing to the development of the draft GPGeneQ questionnaire. Seven GPs (wave 2) partic-

ipated in a group-based questionnaire verification (GBQV) process. Pilot testing of the draft GPGeneQ was conducted on a third wave of participants; 145 GPs of 196 who attended the education program and completed the questionnaire within a week prior to the intervention (participation rate, 74%). Selected demographics and current practice of these participants (wave 3) are described in Table 2. Although these GPs may have been early adopters of genetics education, their demographics were similar to those of the wider Australian population of GPs, except that there was a greater proportion of females in our study population (Table 3). Of these 145 GPs, 125 also completed the assessment questionnaire 1 month after the workshop (86.2%).

Theoretical framework

GPGeneQ was based on the intervention The Genetics File resource and workshop with respect to the topics that were assessed. These topics were deemed to be important and relevant in review of the literature and consultations with GPs, GP educators, and consumers as well as a previous needs assessment (Metcalfe et al. 2002). Both aspects of the intervention were based on ALT principles. Drawing on a model for similar psychometric evaluations (Hutchinson 1999), a program logic model for the intervention was

Table 2 Current practice of GPs participating in pilot testing (total $N=145$)

Characteristic	Number	Valid% ^a
Previous CPD (in which main topic was genetics)		
Yes	35	25.2
No	104	74.8
Referred to		
Genetics clinic	95	66.4
Familial cancer clinic	59	44.4
Self-rated knowledge of genetics		
No knowledge	2	1.4
Very poor	17	11.7
Poor	71	49.0
Satisfactory	55	37.9
Good	0	0.0
Very good	0	0.0
Self-rated comfort in dealing with genetics issues		
Very uncomfortable	0	0.0
Uncomfortable	24	16.6
Neither comfortable nor uncomfortable	81	55.9
Comfortable	34	23.5
Very comfortable	6	4.1

^a There were missing data for some respondents, therefore the percentage is reported as valid%, i.e. the percentage of actual responses to that item

Table 3 Comparison of demographics of GPGeneQ participants (in pilot testing) and population-derived GPs in Australia (total $N=145$)

Characteristic	GPGeneQ GPs		Australian GPs (2003–2004) ^a				
	<i>N</i>	Valid % ^b	<i>N</i>	%	χ^2	<i>df</i>	<i>p</i> value
Age							
<35 years	13	9.2	2,548	11.1	0.69	3	0.875
35–44 years	42	29.8	6,490	28.3			
45–54 years	47	33.3	7,333	32.0			
≥55 years	39	27.7	6,578	28.7			
Gender							
Male	75	51.7	14,577	63.5	8.64	1	0.003
Female	70	48.3	8,372	36.5			
Place of graduation							
Australia	114	78.6	16,506	71.9	3.20	1	0.074
Other	31	21.4	6,443	28.1			

^a Obtained from the Australian Government, Department of Health and Ageing, Primary Care Division at: (<http://www.health.gov.au/internet/main/publishing.nsf/Content/General+Practice+Statistics-1>, accessed August 2010)

^b There were missing data for some respondents, therefore the percentage is reported as valid%, i.e. the percentage of actual responses to that item

developed, which is described elsewhere (Gaff et al. 2007). From this, the framework for the questionnaire was developed to include four sections (demographics and current practice, and scales for knowledge, behaviour and attitude), generating an item bank with a total of 102 items (Table 1).

Content validity

During development of the questionnaire, the item bank was revised several times with major changes to the knowledge and attitude items, whilst there were fewer changes to the behaviour items (Table 1). After round 1 of the Delphi the main modifications were to reduce the number of knowledge items (from 40 to 18) and include an ‘unsure’ response, whilst the attitude items were also reduced in number (from 34 to 27) and modified to include Likert-type responses. The open-ended responses for the behaviour items were generated, ranked and coded. The modified version of the item bank was then used in the second Delphi round. The results (data not shown) after two rounds of the Delphi process indicated that most participants felt that the items were very important and only somewhat difficult for GPs to answer post-educational intervention; hence no further questions were removed. All participants in the second round either agreed or strongly agreed with this and, although further consensus rounds were planned, consensus was achieved after two rounds, thus resulting in the production of the draft GPGeneQ questionnaire.

Ecological validity

During the GBQV, the GPs (five in wave 1 and two in wave 2) generally agreed with the expert panel regarding the relevance, clarity, ambiguity, and difficulty of the items, confirming the results of the Delphi process. GPs generally preferred tick-box-type items, but accepted that a more accurate measure of behaviour would be achieved by not providing options or responses for GPs. Generally GPs felt that the assessment questionnaire was a reasonable length, and could be completed relatively quickly (approximately 15–20 min).

Construct and criterion validity

Following pilot testing, a further 19 items from the draft GPGeneQ were discarded based on the selection criteria described in the methods. Eight knowledge items were discarded. The behaviour and attitude items were subjected to psychometric PCA and reliability analyses and the results summarised in Table 4 for the behaviour and attitude scales. The PCA showed that each item in the behaviour scale, and 16 of the 27 items in the attitude scale had a loading of at

least 0.30, inferring that these 16 items are related to each other and measure the same attitude construct. Eleven of the items on the attitude scale did not load on the same principal component, suggesting that they were measuring a different construct. These items were therefore omitted. The total variance explained by each scale was 34.3% for behaviour, and 56.9% for attitudes. Independence of each scale was shown by the fact that items from the behaviour scale did not load with items from the attitude scale.

Cronbach's α measured internal consistency for both scales, with values of 0.67 for the behaviour scale and 0.79 for the attitude scale. Item-total correlations for each item within a scale were greater than 0.20, hence no further items were removed from either scale. Furthermore, the α value for each scale decreased upon deletion of any one item within the scale.

The final assessment questionnaire: GPGeneQ

The resulting assessment questionnaire, GPGeneQ, comprised all four sections including: (1) demographics and current practice; (2) genetics knowledge; (3) self-reported

Table 4 Psychometric analyses, including principal components and reliability analysis for the behaviour and attitude scales

Item	Principal components analysis			Reliability analysis		
	PC	Eigenvalue	% variance	r_{it}	α if item deleted	α
Behav1	0.60			0.39	0.64	
Behav2	0.56			0.36	0.64	
Behav3	0.65			0.44	0.62	
Behav4	0.57			0.36	0.64	
Behav5	0.65			0.46	0.61	
Behav6	0.53			0.34	0.65	
Behav7	0.52			0.34	0.65	
		2.4	34.3%			0.67
Att1	0.51			0.39	0.78	
Att2	0.34			0.27	0.79	
Att3	0.51			0.42	0.77	
Att4	0.40			0.33	0.78	
Att5	0.65			0.52	0.77	
Att6	0.57			0.46	0.77	
Att7	0.34			0.25	0.79	
Att8	0.58			0.46	0.77	
Att9	0.68			0.58	0.76	
Att10	0.37			0.30	0.78	
Att11	0.48			0.36	0.78	
Att12	0.51			0.40	0.78	
Att13	0.55			0.41	0.77	
Att14	0.55			0.46	0.77	
Att15	0.61			0.49	0.77	
Att16	0.39			0.30	0.78	
		1.01	56.9%			0.79

behaviour; and (4) attitudes. The knowledge scale was made up of ten true/false/unsure items, with scores ranging from -10 to $+10$. The behaviour scale included seven open-ended items, with scores ranging from 0 to 28 . The attitude scale comprised 16 items, with Likert-type response alternatives, and scores ranging from -32 to $+32$. The GP GeneQ can be found in the Electronic supplementary material ([Online resource 1](#)) and the scoring scheme in the Electronic supplementary material ([Online resource 2](#)).

Table 5 shows the results of paired sample t tests for all three scales at two time points from the piloting of the GP GeneQ. These data indicate that all scales were able to detect changes over time (i.e. pre-test vs post-test results), which showed improvements that were significant (knowledge change, $p < 0.001$; CI, -1.63 to -0.68 ; behaviour change, $p < 0.001$; CI, -4.15 to -2.21 ; attitudinal change, $p = 0.002$; CI, -2.68 to -0.62), thereby demonstrating sensitivity of the instrument.

Discussion

In recent years, education strategies have been developed as a means of improving the genetic literacy of health practitioners in a number of countries (e.g. USA, National Coalition for Health Professional Education in Genetics (NCHPEG 2010) and Genetics in Primary Care (Burke et al. 2002); UK, the NHS National Genetics Education and Development Centre (NGEDC 2010); UK and Europe, Eurogentest (EuroGentest 2010). In parallel with program development and delivery, rigorous evaluation is needed to determine the impact of educational programs and provide an evidence base for the development of future interventions (Metcalf et al. 2008).

When evaluating a program, the outcomes measured need to reflect the aims of the program (Metcalf et al. 2008). Assessment measures should be able to answer the evaluation questions asked (i.e. thereby ensuring specificity and content validity); measure change (i.e. addressing

sensitivity as an edumetric measure); and be appropriate for the target group (i.e. ensuring ecological validity). We have described the development and initial validation of a questionnaire to assess GPs' knowledge, behaviour and attitudes that fulfil these criteria.

Knowledge scale

Items were shown to be very important and easily understood by experts in the Delphi process as well as GPs in the initial group-based questionnaire verification. Furthermore, the results indicate a significant difference in the mean total knowledge score following the intervention. While all concepts were covered in the resource, not all were emphasised in the workshop which was focussed on selected topics from the resource, as the workshop time was limited to 2 h, and thus the scale is relatively simple and concise for ease of use. The range of items in the knowledge scale have broad relevance to general practice and so the scale could be applied to other educational interventions relating to genetics in primary care, providing the scale is validated in the new context.

Behaviour scale

PCA demonstrated that the behaviour items formed a uni-dimensional scale. Although only 34.3% of the variance in responses was explained by the items in the scale, all behaviour items loaded well on the same principal component. In addition, the Cronbach's α for internal consistency of this scale met Cortina's standard of $\alpha > 0.6$ for an experimental scale (Cortina 1993), and approached $\alpha > 0.7$ (Nunnally 1978). These limitations may be explained firstly due to the sensitivity of Cronbach's α to the test length. The α value is a function of the number of items in the scale and the correlation between items (Nunnally 1978). The behaviour scale comprised less than ten items; therefore, a more accurate measure of reliability would be the item-total

Table 5 Paired sample t test for the knowledge, behaviour and attitude scales at two time points

Scale	Pre-test (mean $\pm$ SEM)	Post-test (mean $\pm$ SEM)	Mean difference $\pm$ SEM	95% CI of difference	t	df^a	p value*
Knowledge ^b	1.24 $\pm$ 0.22	2.40 $\pm$ 0.26	-1.15 $\pm$ 0.24	-1.63 to -0.68	-4.77	123	<0.001
Behaviour ^c	16.77 $\pm$ 0.42	19.95 $\pm$ 0.41	-3.18 $\pm$ 0.49	-4.15 to -2.21	-6.50	120	<0.001
Attitude ^d	16.81 $\pm$ 0.55	18.46 $\pm$ 0.57	-1.65 $\pm$ 0.52	-2.68 to -0.62	-3.18	122	0.002

SEM standard error of the mean, CI confidence intervals, df degrees of freedom

* $p < 0.05$, significance

^a Although 145 GPs completed the first questionnaire before the intervention, only 125 completed both (pre-test/post-test). Also, there are missing values since this is a paired sample t test, as not every participant completed each scale

^b Knowledge scale ranges from -10 to $+10$

^c Behaviour scale ranges from 0 to 28

^d Attitude scale ranges from -32 to $+32$

correlation coefficients, which are within the accepted limits (Briggs and Cheek 1986). Secondly, behaviour items were open-ended scenarios, which were coded based on the appropriateness of each response, as decided by the experts during the Delphi process. These items are therefore disparate, resulting in a decreased correlation between items. It should be noted that, although these are open-ended items, there are standards for the coding of responses, meaning these are items with undisclosed scales for scoring. Although unusual, these have the advantage of ensuring no information is missed, which may be the case with closed items (such as multiple choice questions), and avoid prompting responses. To our knowledge, this is the first example of a behaviour scale that uses pre-coded open-ended items to measure the effects of an education intervention in genetics. In different healthcare settings and other countries, the responses could be coded differently based on that which is considered appropriate behaviour for that context.

Attitude scale

Principal component values for all attitudinal items were greater than 0.30, suggesting the items form an independent and reliable attitude scale ($\alpha=0.79$). Although there is some discussion in the literature suggesting that loadings of less than 0.40 are not as well correlated as those over 0.40, we used Rummel's criterion (>0.30) (Rummel 1970). When only one factor is being investigated, this loading value is equally stringent as 0.40, and allows for greater diversity of the issue at hand, i.e. greater diversity in the range of attitudes assessed.

Sensitivity of the questionnaire

As all three scales were able to detect significant differences over time for the GPs in the pilot testing, the instrument is sensitive to measuring change and therefore provides evidence for the validity of the questionnaire.

Application of the questionnaire

GP learning has focussed for many years on improving GPs' competencies and skills (knowledge, behaviour and attitudes) and a rigorous assessment of most GP education interventions would be expected to include a validated instrument that measures these three components of competencies. Since the measures used in GPGenQ are fairly generic they may be transferable to other genetics educational settings with GPs and possibly with other health professionals in primary care, with the proviso that it should not be assumed that the questionnaire will be specific for a new target audience; as with any measurement instrument, it should ideally be tested again for ecological validity with the specific group of health professionals and modified as necessary.

Expectations of education interventions have progressed to knowledge translation (exchange, synthesis and ethical application of knowledge) (Davis et al. 2003), with educational interventions likely to be complex in design. As noted by Davis and colleagues, at the base of knowledge translation is predisposition to change which is brought about through increasing knowledge or skills. Regardless of how genetics knowledge and clinical skills in primary practice are disseminated, there is a need to evaluate the effectiveness of such dissemination in order to demonstrate that changes in knowledge and clinical skills have taken place. This assessment questionnaire was developed to evaluate a multifaceted and complex learning intervention in genetics, as part of a systematic effort to improve genetics education in primary care in Australia, and result in behavioural change, as well as improvement in knowledge and attitudes. More detailed longitudinal findings on a larger population of GPs who received this intervention will be reported elsewhere. However, the authors acknowledge that additional outcome measures are required to capture knowledge translation, and it can be debated to what extent data collection by questionnaire alone can evaluate this.

Conclusion

In summary, the current results provide evidence for validity and reliability of the GPGenQ, specifically developed to measure the impact of educational interventions on self-reported genetics practice by GPs. The questionnaire may have broader application in education of other non-genetics health professionals, although with need for further testing and refinement for each specific setting. While the questionnaire was developed and applied several years ago, the content and context are still currently relevant, and the procedures involved have not been previously articulated for a genetics audience. We have described this process of developing and validating the components of the questionnaire in sufficient detail to allow others in genetics education to consider using the instrument *per se*, or to adapt it for their specific target audience and setting. The GPGenQ questionnaire and the coding scheme are also available on request.

Acknowledgements The authors are grateful to the Victorian Department of Human Services, Biotechnology Australia (a federal government initiative) and the Cooperative Research Centre for the Discovery of Genes for Common Human Diseases for financial support. Anna Flouris received financial support through an Australian Postgraduate Award. The authors declare they have no conflict of interest. We are also grateful to the Delphi process panel of experts, Michael Batchelor, Jane Wallace, Lydia Gaffney, the many genetic counsellors and GP facilitators who helped with the workshops, and the GPs who attended the workshops and completed the questionnaires.

Conflicts of interest The authors declare that they have no conflict of interest.

References

- Baars MJ, Henneman L, Ten Kate LP (2005) Deficiency of knowledge of genetics and genetic tests among general practitioners, gynecologists, and pediatricians: a global problem. *Genet Med* 7(9):605–610
- Boynton PM, Greenhalgh T (2004) Selecting, designing, and developing your questionnaire. *BMJ* 328(7451):1312–1315
- Briggs SR, Cheek JM (1986) The role of factor analysis in the development and evaluation of personality scales. *J Personality* 54:106–148
- Burke W, Emery J (2002) Genetics education for primary-care providers. *Nat Rev Genet* 3(7):561–566
- Burke W, Acheson L, Botkin J, Bridges K, Davis A, Evans J, Frias J, Hanson J, Kahn N, Kahn R, Lanier D, Pinsky LE, Press N, Lloyd-Puryear MA, Rich E, Stevens N, Thomson E, Wartman S, Wilson M (2002) Genetics in primary care: a USA faculty development initiative. *Community Genet* 5(2):138–146
- Carroll JC, Rideout AL, Wilson BJ, Allanson JM, Blaine SM, Esplen MJ, Farrell SA, Graham GE, MacKenzie J, Meschino W, Miller F, Prakash P, Shuman C, Summers A, Taylor S (2009) Genetic education for primary care providers: improving attitudes, knowledge, and confidence. *Can Fam Phys* 55(12):e92–e99
- Carver RP (1974) Two dimensions of tests: psychometric and edumetric. *Am Psych* 29:512–518
- Clyman JC, Nazir F, Tarolli S, Black E, Lombardi RQ, Higgins JJ (2007) The impact of a genetics education program on physicians' knowledge and genetic counseling referral patterns. *Med Teach* 29(6):e143–e150
- Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. *J Appl Psych* 78(1):98–104
- Davis D, Evans M, Jadad A, Perrier L, Rath D, Ryan D, Sibbald G, Straus S, Rappolt S, Wowk M, Zwarenstein M (2003) The case for knowledge translation: shortening the journey from evidence to effect. *BMJ* 327(7405):33–35
- De Vaus DA (1995) Surveys in social research (Chapter 6). Allen & Unwin, North Sydney
- Emery J, Watson E, Rose P, Andermann A (1999) A systematic review of the literature exploring the role of primary care in genetic services. *Fam Pract* 16(4):426–445
- EuroGentest (2010). Available from: <http://www.eurogentest.org/professionals/education/>. Accessed August 2010.
- Gaff CL, Aitken M, Flouris A, Metcalfe SA (2007) A model for the development of genetics education programs for health professionals. *Genet Med* 9(7):451–457
- Greendale K, Pyeritz RE (2001) Empowering primary care health professionals in medical genetics: how soon? How fast? How far? *Am J Med Genet* 106(3):223–232
- Guadagnoli E, Velicer WF (1988) Relation of sample size to the stability of component patterns. *Psychol Bull* 103(2):265–275
- Hofman KJ, Tambor ES, Chase GA, Geller G, Faden RR, Holtzman NA (1993) Physicians' knowledge of genetics and genetic tests. *Acad Med* 68(8):625–632
- Hutchinson L (1999) Evaluating and researching the effectiveness of educational interventions. *BMJ* 318(7193):1267–1269
- Kaufman DM (2003) Applying educational theory in practice. *BMJ* 326(7382):213–216
- Knowles MS (1984) Applications in continuing education for the health professions. *Andragogy in action*. Jossey-Bass, San Francisco, pp 297–340
- Kolb SE, Aguilar MC, Dinenberg M, Kaye CI (1999) Genetics education for primary care providers in community health settings. *J Community Health* 24(1):45–59
- Kumar S, Gantley M (1999) Tensions between policy makers and general practitioners in implementing new genetics: grounded theory interview study. *BMJ* 319(7222):1410–1413
- Linstone HA, Turoff M (1975) The Delphi method: techniques and applications. Addison-Wesley Pub. Co. Advanced Book Program, Reading
- McLeod P, Steinert Y, Meterisian S, Child S (2004) Using the Delphi process to identify the curriculum. *Med Educ* 38(5):548
- Metcalfe S, Hurworth R, Newstead J, Robins R (2002) Needs assessment study of genetics education for general practitioners in Australia. *Genet Med* 4(2):71–77
- Metcalfe S, Seipolt M, Aitken M, Flouris A (2005) Educating general practitioners about prenatal testing: approaches and challenges. *Prenat Diagn* 25(7):592–601
- Metcalfe SA, Aitken M, Gaff CL (2008) The importance of program evaluation: how can it be applied to diverse genetics educational settings? *J Genet Couns* 17(2):170–179
- NCHPEG (2010) National Coalition for Health Professional Education in Genetics. Available from: <http://www.nchpeg.org/>. Accessed August 2010.
- Newble DI (2002) Don't presume about experienced adult learners in medicine. *BMJ* 325(7367):779
- Newman P, Peile E (2002) Valuing learners' experience and supporting further growth: educational models to help experienced adult learners in medicine. *BMJ* 325(7357):200–202
- NGEDC (2010) National Genetics Education and Development Centre. Available from: <http://www.geneticseducation.nhs.uk/>. Accessed August 2010.
- Nunnally JC (1978) Psychometric theory. McGraw-Hill, New York
- RACGP (2005) Royal Australian College of General Practitioners: a quality framework for Australian General Practice: background paper
- RACGP (2007) Royal Australian College of General Practitioners: curriculum for Australian general practice: learning life of general practitioners. Available from: <http://www.racgp.org.au/curriculum>. Accessed August 2010.
- Robins R, Metcalfe S (2004) Integrating genetics as practices of primary care. *Soc Sci Med* 59(2):223–233
- Rummel R (1970) Applied factor analysis. Northwestern University Press, Evanston
- Streiner DL, Norman GR (1995) Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford
- Sullivan FM, Lockhart PM, Usherwood TP (2010) Potential implications of genomic medicine in general practice. *Med J Aust* 193(2):120–123
- Suther S, Goodson P (2003) Barriers to the provision of genetic services by primary care physicians: a systematic review of the literature. *Genet Med* 5(2):70–76
- Tabachnick BG, Fidell LS (2001) Using multivariate statistics (xxvi, 96 discussions). Allyn and Bacon, Boston, MA