

Developing and testing items for the South African Personality Inventory (SAPI)

Authors:

Carin Hill¹
Jan Alewyn Nel²
Fons J.R. van de Vijver^{2,3,4}
Deon Meiring^{5,6}
Velichko H. Valchev^{3,5}
Byron G. Adams^{1,3}
Gideon P. de Bruin¹

Affiliations:

¹Department of Industrial Psychology and People Management, University of Johannesburg, South Africa

²School of Human Resource Sciences, North-West University, South Africa

³Department of Cross-Cultural Psychology, Tilburg University, The Netherlands

⁴School of Psychology, University of Queensland, Australia

⁵Department of Human Resource Management, University of Pretoria, South Africa

⁶Department of Industrial Psychology, Stellenbosch University, South Africa

Correspondence to:
Carin Hill

Email:
chill@uj.ac.za

Postal address:
PO Box 524, Auckland Park
2006, South Africa

Dates:
Received: 09 Apr. 2013
Accepted: 18 July 2013
Published: 12 Nov. 2013

How to cite this article:
Hill, C., Nel, J.A., Van de Vijver, F.J.R., Meiring, D., Valchev, V.H., Adams, B.G. *et al.* (2013). Developing and testing items for the South African Personality Inventory (SAPI). *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 39(1), Art. #1122, 13 pages. <http://dx.doi.org/10.4102/sajip.v39i1.1122>

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Orientation: A multicultural country like South Africa needs fair cross-cultural psychometric instruments.

Research purpose: This article reports on the process of identifying items for, and provides a quantitative evaluation of, the South African Personality Inventory (SAPI) items.

Motivation for the study: The study intended to develop an indigenous and psychometrically sound personality instrument that adheres to the requirements of South African legislation and excludes cultural bias.

Research design, approach and method: The authors used a cross-sectional design. They measured the nine SAPI clusters identified in the qualitative stage of the SAPI project in 11 separate quantitative studies. Convenience sampling yielded 6735 participants. Statistical analysis focused on the construct validity and reliability of items. The authors eliminated items that showed poor performance, based on common psychometric criteria, and selected the best performing items to form part of the final version of the SAPI.

Main findings: The authors developed 2573 items from the nine SAPI clusters. Of these, 2268 items were valid and reliable representations of the SAPI facets.

Practical/managerial implications: The authors developed a large item pool. It measures personality in South Africa. Researchers can refine it for the SAPI. Furthermore, the project illustrates an approach that researchers can use in projects that aim to develop culturally-informed psychological measures.

Contribution/value-add: Personality assessment is important for recruiting, selecting and developing employees. This study contributes to the current knowledge about the early processes researchers follow when they develop a personality instrument that measures personality fairly in different cultural groups, as the SAPI does.

Introduction

Problem statement

In South Africa, personality research and assessment has been gaining momentum in the last 10 years (Meiring, Van de Vijver, Rothmann, & Barrick, 2005; Taylor, 2000; Visser & Viviers, 2010). The need for measures that are more sensitive to ethnic differences has heightened the interest in personality assessment (Nel *et al.*, 2012; Valchev *et al.*, 2011). This makes fair and comparable measurement more challenging.

South African legislation also provides a legal framework for developing psychological tests. Section 8 of the *Employment Equity Act (Act 55 of 1998)* (Government Gazette, 1998) clearly states that all psychological measures in South Africa should measure concepts fairly and equally for all ethnic groups and deal with cultural, linguistic and racial elements without introducing bias against any group. Differences in culture, the distribution of socioeconomic resources, education and employment statuses in the country further contributes to the challenge of fair psychological measurement in South Africa (Foxcroft & Roodt, 2009). This is particularly relevant in instances where organisations use psychological measures for recruiting, selecting, placing and developing employees.

In an effort to overcome these problems, personality researchers from South Africa and the Netherlands initiated a project eight years ago. This was the South African Personality Inventory project¹ (Meiring, Van de Vijver & Rothmann, 2006; Nel *et al.*, 2012; Valchev *et al.*, 2011; Valchev

1. The SAPI, an acronym for the South African Personality Inventory, is a project that aims to develop an indigenous personality measure for all 11 official languages in South Africa. The participants are Byron Adams (University of Johannesburg and Tilburg University, The Netherlands), Deon de Bruin (University of Johannesburg), Karina de Bruin (University of the Free State), Carin Hill (University of Johannesburg), Leon Jackson (North-West University), Deon Meiring (University of Pretoria and University of Stellenbosch), Jan Alewyn Nel (North-West University), Ian Rothmann (North-West University), Michael Temane (North-West University), Velichko Valchev (Tilburg University, The Netherlands), and Fons van de Vijver (North-West University and Tilburg University, The Netherlands).

et al., 2013). The South African Personality Inventory (SAPI) started with an investigation of indigenous conceptions of personality in the 11 official South African languages.

The project initially used a combined etic-emic approach in which native speakers of each of the 11 languages described persons they know well (like parents and friends) and did not know well (teachers or neighbours). After content analysis of these responses, the researchers identified unique facets (specific to certain languages) and common facets (that most or all languages shared) (Nel *et al.*, 2012; Valchev *et al.*, 2011; Valchev *et al.*, 2013).

The researchers clustered these facets further. This resulted in nine general personality clusters (extraversion, soft-heartedness, conscientiousness, emotional stability, intellect, openness, integrity, relationship harmony and facilitating). The personality structure that the researchers derived comprised a three-tiered hierarchical structure. It had nine clusters at the top, 37 sub clusters (between two and six sub clusters per cluster) and 190 personality facets at the lowest level.

Three of these nine clusters corresponded with the big five factors (extraversion, conscientiousness and emotional stability) (see John & Srivastava, 1999), two were reminiscent of the openness factor of the big five, (openness and intellect) and one was similar to the Honesty factor in the HEXACO model (integrity) of Ashton and Lee (2007). Three SAPI clusters, namely soft-heartedness, relationship harmony and facilitating, were unique to the SAPI although some elements of the big five's agreeableness factor were present in these unique clusters.

The overall objectives of this article are to discuss the phase of the project in which the authors developed items using the responses they obtained during the qualitative phase and to investigate the construct validity and reliability of the instruments they developed to measure the nine clusters.

Because the SAPI is still in its early development stage, the authors determined construct validity only to construe which items performed well within a specific cluster. According to Bayoglu, Unal, Elibol, Karabulut and Innocenti (2013), researchers should first determine reliability and construct validity for a new questionnaire before attempting further validity studies. This information pertains to developing and testing items.

Literature review

Developing the South African Personality Inventory item pool

According to Saucier (2008), the way one defines personality is significant because it affects how researchers will select variables when they study personality.

When researchers define personality constructs in different cultural contexts, they usually use a top-down approach (the etic approach) because they investigate and test the generalisability of Western models and theories of personality

in non-Western cultural contexts (Cheung, Van de Vijver & Leong, 2011). This approach assumes that the personality traits that these Western models and theories measure adequately represent personality dimensions in other cultures. However, Cheung *et al.* (2011, p. 595) discussed various substantive and methodological issues related to the etic approach, noting that researchers might miss: '... indigenous constructs that are salient in the local folk concepts of personality and in the local taxonomy of person descriptions' when they use imposed etic measures.

The aim of indigenous psychology (the emic approach) is to develop an insider's perspective of psychological phenomena in a culture. Using conceptions and methodologies that are embedded in, and derived from, the ethnic or cultural group being studied to generate knowledge and enrich mainstream psychology typifies this approach (Chui, Kim, & Wan, 2008; Ho, 1998; Ho, Peng, Lai, & Chan, 2001).

Cheung *et al.* (2011) contended that researchers need combined etic and emic approaches in order to enlarge current conceptualisations of universal personality constructs, thereby possibly bridging the gap between mainstream and indigenous psychology and demarcating the universal and culturally specific aspects of psychological constructs. This combined etic-emic approach can include:

- ... (a) the use of a combination of etic and emic measurement, (b) studies in which universal and culture-specific aspects are delineated in an iterative process of data collections with continually adapted instruments, and (c) the use of mixed methods (e.g., the use of an etic measure combined with interviews for collecting information about culture-specific features not covered by the etic instrument). (Cheung *et al.*, 2011, p. 597)

For a complete description of this convergent approach, where researchers combine the etic and emic approaches, see the work of Cheung *et al.* 2011; Nel *et al.* 2012.

In this project, the SAPI research team used a combined etic-emic approach to identify culturally and linguistically adequate personality descriptive terms for all 11 official South African languages (Cheung *et al.*, 2011; Nel *et al.*, 2012). Current etic models, like the Five Factor Model (FFM) and the HEXACO model (Nel *et al.*, 2012), partly informed the clustering of these emic terms.

The next step in the project was to use construct modelling to develop a culturally appropriate measuring instrument using these personality descriptive terms.

Construct modelling is a framework for developing a measuring instrument based on using four building blocks:

1. Explaining the construct to others with the help of the construct map.
2. Creating items believed to lead respondents to give responses that will indicate levels of the construct map.
3. Trying out those items with a sample of respondents.
4. Analysing the resulting data to check whether the results are consistent with the initial intentions as expressed in the construct map (Wilson, 2005).

Taking a combined etic-emic approach, an adapted version of Wilson's (2005) proposed construct map, the authors used a modelling framework to develop the item pool for each of the identified SAPI clusters. The process involved the following steps:

1. Grouping the original responses by facet and extracting content-representative responses.
2. Describing the nine SAPI constructs (i.e. also defining these constructs) with the help of the content-representative responses.
3. Transforming qualitative responses into item stems.
4. Designing items the authors believed tapped the facets.
5. Administering the items to a sample of respondents.
6. Analysing the resulting data to check whether the results are consistent with the initial intentions as stated in the construct map.

A detailed discussion of each of these building blocks, in relation to developing the SAPI items, follows.

Grouping the original responses and extracting content-representative responses

In order to extract content representative responses, the authors first grouped all the original responses associated with each SAPI cluster according to language (for example, extraversion clustered according to Afrikaans, English, Swati, Ndebele, IsiXhosa and IsiZulu). They then grouped the responses from the 11 languages as they related to the various sub clusters (for example, all the responses from all the languages for emotional stability clustered according to balance, courage and ego strength). The authors then clustered the facets in each sub cluster as the various language groups presented them. Finally, the authors examined the original responses by facet and retained only those that represented the facets as content-representative responses.

The authors used these content-representative responses to develop the construct maps of the SAPI.

The construct maps of the South African Personality Inventory

One can define a construct map as a one-dimensional latent variable. However, because the SAPI structure is multidimensional, Wilson (2005) suggested dealing with one dimension at a time. This would make it possible to use the construct modelling framework for developing the SAPI. A construct map contains a coherent and substantive definition of the content of the construct as well as an idea that the construct comprises an underlying continuum (Wilson, 2005). Similarly, Boyle, Matthews and Saklofske (2008) stated that the key to moving forward with psychometrically sound measurement rests with the definitions that researchers decide best represents a particular domain of behaviour, psychological disturbance (or wellbeing) or underlying traits, like extraversion or neuroticism.

The authors took into account that, in addition to being multidimensional, the SAPI is also hierarchical in its structure. It consists of clusters, sub clusters, facets and items.

With this in mind, the authors developed definitions for the clusters and facets in the SAPI. For a complete description of the definitions of the SAPI clusters, see Nel *et al.* 2012.

The authors developed the definitions of the various facets using the shared content in the content-representative responses. They discarded facets that consisted of four or fewer descriptions and those that fewer than two language groups presented (for example, the authors discarded dutiful and deliberating from conscientiousness: talented and useless from intellect; prim and proper from openness; and satisfying others and wrathful from soft-heartedness).

During this process, three new facets emerged. These were gullible (in soft-heartedness), shamed (in emotional stability) and impulsive (in emotional stability). the authors divided the original introvert or extravert facet (in extraversion) into two separate facets and developed definitions for the remaining 184 facets.

Table 1 contains examples of definitions for each of the remaining facets.

Item stems

The next step involved transforming the content-representative responses into item stems (see Table 2 for an example and Nel *et al.* [2012] for examples of the original responses). The authors then used these item stems in the next building step: the process of designing the items.

The process of designing items

The process of designing items consisted of four parts:

1. developing stimuli or items to which the participants respond
2. deciding on a response format or method
3. determining conditions that govern how participants respond to the stimuli
4. establishing procedures for scoring the responses (Hogan, 2007).

For the first part of the process, the authors rephrased the SAPI item stems as items that they could use in the final versions of the questionnaires. The development of the SAPI items followed the psycho-lexical tradition of item development closely. This is an approach that researchers use widely when they examine personality (see Cheung *et al.*, 2011). Researchers who work in the psycho-lexical tradition usually begin their research into personality descriptors in a particular language group with the analysis of the group's dictionary. This resulted in the extraction of a comprehensive collection of personality-descriptive terms that they were subsequently reduce according to a number of criteria (Angleitner, Ostendorf, & John, 1990).

The SAPI project used an adapted version of the psycho-lexical approach. Instead of using South African dictionaries for all the 11 languages, the SAPI project derived its personality descriptors from the qualitative phase's content-representative responses (i.e. the transcripts of the

TABLE 1: Response-derived definitions of the nine South African Personality Index clusters.

Cluster	Sub cluster	Facet	Definition
Conscientiousness	Achievement Orientated	Competitive	Like to compete with others; challenge others. Enjoying being in competition.
	Dedication	Perseverant	Persevering and not giving up in the face of hardships or difficulty. Facing challenges successfully and putting problems behind oneself. Continuing with what one is doing until one has completed it.
	Orderliness	Punctual	Doing things on time and expecting from others to be on time. Sticking to allocated time, keeping track of time and doing things according to a time schedule.
	Self-discipline	Rebellious	Not conforming to traditional customs and values; having a contrarian attitude, being 'against' things in principle. Making one's own rules and regulations.
	Thoughtless	Reckless	Not being cautious and careful, not considering possible dangers and consequences.
Emotional stability	Balance	Balancing Life	Living a balanced life by dedicating equal amounts of time to family, health and work or study needs.
	Courage	Fearful	Being afraid of people and situations, being worried about one's health and one's own and other people's safety.
	Ego-strength	Needy	Trying to please people to achieve acceptance, recognition, love and praise. Finding it difficult to go somewhere alone.
	Emotional Control	Temperamental	Acting in an unpredictable way, being easily angered and provoked, treating others badly for no reason and having frequent mood swings.
	Emotional Sensitivity	Emotional	Showing feelings like anger, excitement, sadness and happiness through crying, facial expressions and body language.
	Neuroticism	Depressive	Feeling down, based on one's life or work situation, disliking oneself and lacking direction and interest in life.
Extraversion	Dominance	Assertive	Not allowing others to manipulate or control one's actions. Ensuring that everyone is aware of the boundaries that one has set and communicating one's views about the behaviour of others.
	Expressiveness	Outspoken	The inclination to share one's feelings or problems with others as a way of sharing burdens and being frank and forthright about one's feelings in a non-threatening manner.
	Sociability	Extrovert	Enjoying the company, companionship and presence of other people, being loud, sociable and expressive and being able to connect with others easily.
Facilitating	Guidance	Leading	Having leadership qualities; have and like to take the role of a leader in one's close relations and in one's broader social environment; enjoying leading.
	Encouraging Others	Having aspirations for others	Hoping for the best for others and that they will succeed in life. Like to see other people progress and realise their potential. Promote others' success and well-being. Having dreams and ambitions for others and wishing others well.
Integrity	Integrity	Morally Conscious	Have good morals and values, a sense of what is right or wrong and being righteous.
	Fairness	Discriminative	Being prejudiced towards others with different orientations, backgrounds, and beliefs.
Intellect	Aesthetics	Artistic	The inclination, likings or preference for engaging or appreciating the arts.
	Reasoning	Intellect	Being bright, shrewd, informed and generating good ideas. Being able to understand concepts easily and learn quickly. Contemplate the content of questions in order to give precise and accurate answers in line with expectations.
	Skilfulness	Enterprising	Taking the initiative to exploit new business opportunities, having business knowledge and the creativity to make and sell items.
	Social Intellect	Perceptive	Being observant, reading the social environments for cues and having insight about others.
Openness	Broadmindedness	Dreamer	Having dreams; dreaming of the future and of things one would like to see in the future; dreaming about others; pursuing one's dreams.
	Epistemic Curiosity	Eager to learn	Being eager, willing and determined to learn, showing an interest in schoolwork and school related academic activities.
	Materialism	Materialistic	Being materialistic and concerned about wealth; spending money on expensive articles.
	Openness to experience	Adventurous	Being an adventurous person who enjoys exploring things to gain new experiences.
Relationship harmony	Approachability	Accommodating	Knowing how to deal with others, making others feel at home and explaining things when others do not understand them.
	Interpersonal Relatedness	Forgiving	Accepting apologies from others, not holding grudges, not taking revenge and preferring to make peace by talking about things.
	Conflict-seeking	Troublesome	Causing trouble, arguments, fights and conflict between people.
	Meddlesomeness	Interfering	Meddling and getting involved in other people's affairs when one's involvement is not needed, collecting information about other people's affairs that do not concern one, intruding and telling others what to do when it is none of one's business, invading others' privacy and prying into their affairs.
Soft-heartedness	Gratefulness	Appreciative	Expressing appreciation, acceptance and love. Liking, adoring, enjoying, and being fond of objects, persons, and/or situations in general.
	Active Support	Community Involvement	Taking interest in, caring for, or serving the community or its development; gaining respect by being a role model for the community.
	Hostility	Critical	Being critical and insulting towards others and opposing others' lifestyles. Being outspoken and looking for weaknesses in others.
	Amiability	Kind	Having a soft heart and being soft-spoken. Being gentle and nice towards people.
	Egoism	Greedy	Wanting more than what one already has. Not being satisfied with what one has.
	Empathy	Caring	Being aware of others' needs and doing things for others. Looking after others and being concerned about the welfare of others. Having others' best interests at heart, being interested in others.

TABLE 2: Developing item stems.

Cluster	Sub cluster	Facet	Original response	Content-representing response	Item stem
Intellect	Social Intellect	Perceptive	She could easily see when you had a problem (IsiZulu)	See when someone has a problem	Seeing when others have problems
Relationship Harmony	Approachability	Accommodating	Addresses us in English so we could understand (IsiXhosa response)	Addresses us in English so we could understand	Addressing others in mutually understandable language

utterances about personality the authors collected from interviews in the first qualitative stage of the SAPI project). They then transformed them into item stems. Whilst they were converting the item stems into items, they followed these general guidelines for writing items to ensure their standardisation:

1. Items had to be short, simple and clear.
2. Items were written in the first person, starting with 'I' followed by concrete behaviours, objects and contexts.
3. Negatives should not be used in the main parts of items.
4. Items that described a single activity or habit were avoided.
5. Temporal qualifiers, like often, always and sometimes were avoided.
6. Items had to be formulated in the direction of the construct.
7. Double-barrelled items were not allowed.
8. Items had to refer to concrete behaviours and not beliefs, values or orientations.
9. Psychological trait terms had to be avoided.
10. Items should not use idioms and expressions or sayings in order to avoid confusion.
11. Items had to be written in English so that they could be translated (cf. Hendriks, Hofstee, & De Raad, 1999).

It is a general rule of thumb, when one develops psychological instruments, to develop two to three times as many items as one needs for the final instrument initially (Hogan, 2007). During the process of developing items for the SAPI, the authors developed 2573 items for the nine clusters:

- 240 items for conscientiousness
- 325 items for emotional stability
- 379 items for extraversion
- 184 items for facilitating
- 117 items for integrity
- 202 items for intellect
- 244 items for openness
- 400 items for relationship harmony
- 482 items for soft-heartedness

Throughout the process, the authors revisited the qualitative data in order to ensure that the items derived from the original responses were relevant. They presented these items to cultural and language experts (representative of all 11 official languages in South Africa) in order to identify:

1. Items that were not translatable to a language other than English
2. Items that were ambiguous or not understandable to speakers of a particular language
3. Items that were not culturally appropriate for a certain language group.

The choice of item format (part two of the item design process) links to the nature of the construct one is measuring and to practical considerations (Foxcroft & Roodt, 2009).

Researchers usually design personality inventories for administration individually or to groups. Therefore, these instruments need to be easy to score and usually use the format of the popular fixed-response or a Likert-type response format (Foxcroft & Roodt, 2009; Hogan, 2007; Wilson, 2005). The authors chose the five-point Likert-type response format

for the SAPI. Responses ranged from 1 (strongly disagree) to 5 (strongly agree).

The next part of the process of designing items involved determining the conditions that govern how participants respond to the stimulus. Consistent with common practice in personality assessment, participants had unlimited time to complete the questionnaire. The authors designed an answer sheet to accompany the test booklets. This answer sheet can be hand or computer scored.

Finally, establishing procedures for scoring the response is an important part of developing a measuring instrument. When analysing Likert-type responses, researchers treat responses as belonging to a numerical scale. They then either add the items or carry out a factor or latent variable analysis. This results in a factor score that measures a common characteristic of the item the researchers set for a respondent (Dittrich, Francis, Hatzinger, & Katzenbeisser, 2007). In this instance, developing the SAPI scale involved 'summing Likert-type items to yield a score that represents the degree to which the construct being measured is present in the respondent' (Fitzpatrick *et al.*, 2004, p. 334).

The last two building blocks in the construct modelling framework relate to testing the developed items in a sample of respondents and analysing the results to determine whether the items achieved the initial purpose for which the researchers developed them. Researchers achieve this by investigating the items' reliability and validity.

The rest of this article addresses this aspect.

Research design

Research approach

The authors used a quantitative, exploratory research design. The main aim of exploratory research is to formulate and clarify ideas and to develop questions and hypotheses for future research (Struwig & Stead, 2001). In this study, the authors used a survey method for collecting data. They used a cross-sectional survey design because they drew a sample from the population at a specific point in time (Shaughnessy & Zechmeister, 1997). The authors used an exploratory approach to analyse the primary data.

Research method

Sampling and research participants

The authors used a convenience, non-probability sampling frame to achieve the objectives of the study. Convenience sampling is the most commonly used sampling technique in personality research (Van de Vijver & Leung, 2000). The authors recruited the current study's participants ($N = 6735$) from various tertiary education and training institutions in South Africa. Researchers have used this sampling technique before in similar studies (see Avdeyeva & Church, 2005; Cheung *et al.*, 2008; De Raad, Sullot, & Barelds, 2008; Di Blas, 2005; Goldberg & Somer, 2000; Hendriks *et al.*, 1999).

Table 3 shows that most participants were between 18 and 21 years old (67%) and had completed Grade 12 (83%). Most were female (59%) and 94% of the participants rated their English reading ability as good to very good. Participants spoke Afrikaans (30%), English (16%) and IsiZulu (13%) as home languages. The sample comprised African (56%) and White (36%) participants.

Measuring instruments

Developing and refining the items: The immediate objective of this stage was to develop an item pool to measure the facets, sub clusters and clusters using the construct modelling framework that Wilson (2005) described. The authors grouped

the original responses the researchers obtained from the interviews in the qualitative phase of the SAPI. They then extracted content-representative responses and developed definitions for the various facets. They generated item stems using the facets' definitions and the content-representative responses.

The authors developed 2573 items for the nine proposed clusters of the SAPI and assembled them into separate questionnaires for administration to the various populations.

The authors encountered a few challenges in constructing and refining the items before they developed the 2573 items.

TABLE 3: Characteristics of the participants (*N* = 6735).

Item	Category	CN % (<i>n</i> = 1042)	ES % (<i>n</i> = 588)	EX % (<i>n</i> = 415)	FC % (<i>n</i> = 439)	IT % (<i>n</i> = 1023)	IL % (<i>n</i> = 524)	OP % (<i>n</i> = 730)	RH 1 % (<i>n</i> = 507)	RH 2 % (<i>n</i> = 478)	SH 1 % (<i>n</i> = 513)	SH 2 % (<i>n</i> = 476)	Total % (<i>N</i> = 6735)
Age (years)	18–19	1.0	42.3	62.4	62.4	1.1	52.7	40.7	49.9	52.7	54.9	56.2	43.3
	20–21	7.0	41.3	24.1	26.9	7.3	15.5	42.6	27.2	25.7	21.5	20.8	23.6
	22–23	15.7	8.8	4.3	3.4	15.7	13.9	9.7	4.9	7.7	8.7	8.6	9.2
	24–25	21.6	3.4	0.7	1.1	21.1	3.1	2.9	1.4	1.5	1.2	1.4	5.4
	26–27	19.2	1.5	1.2	0.9	19.7	0.8	1.4	0.4	0.2	0.8	0.4	4.2
	28–29	16.7	0.2	0.2	0.2	16.4	0.0	0.3	0.2	0.0	0.4	0.0	3.1
	30–31	12.2	0.9	0.2	0.2	12.3	0.0	1.0	0.2	0.0	0.0	0.4	2.5
	Other	6.3	1.4	5.3	4.3	6.3	14.1	1.4	15.0	12.1	0.8	0.2	6.1
Education	Missing	0.2	0.2	1.6	0.5	0.0	0.0	0.1	0.0	0.0	12.3	12.0	2.5
	Grade 9	0.0	0.0	0.0	0.0	0.0	2.9	0.0	2.0	4.0	1.6	4.0	1.3
	Grade 12	75.5	90.1	87.7	90.4	73.5	83.8	89.0	81.9	79.9	82.7	80.5	83.2
	Certificate	15.3	3.2	2.2	2.1	17.2	2.5	4.1	3.9	3.1	2.5	1.9	5.3
	Diploma	7.7	2.2	2.2	1.8	7.5	2.3	2.1	2.0	2.3	2.7	2.5	3.2
	Bachelors	0.6	3.2	3.4	2.7	0.6	1.9	4.2	1.8	2.9	1.8	2.7	2.3
	Honours	0.1	0.2	0.5	0.7	0.0	0.0	0.3	0.6	0.4	0.4	0.2	0.3
	Master's	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Gender	Doctorate	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
	Other	0.4	0.9	2.4	2.3	0.7	2.3	0.3	3.0	2.3	1.6	2.1	1.7
	Missing	0.5	0.0	1.4	0.0	0.5	4.1	0.0	4.8	4.9	6.7	6.1	2.6
	Male	46.5	20.7	36.9	35.8	48.0	48.3	22.7	42.2	45.2	45.4	47.9	40.0
	Female	53.4	79.3	61.9	64.2	52.0	50.2	77.3	56.4	54.0	53.2	50.8	59.3
	Missing	0.1	0.0	1.2	0.0	0.0	1.5	0.0	1.4	0.8	1.4	1.3	0.7
English Reading Ability	Very poor	0.5	0.7	0.0	0.0	0.4	1.0	0.4	1.0	1.0	0.2	0.2	0.5
	Poor	9.5	1.0	1.0	1.1	8.6	4.8	1.6	3.7	4.6	4.3	4.8	4.1
	Good	11.8	42.0	35.9	36.7	12.3	63.5	35.5	62.7	63.2	63.4	61.1	44.4
	Very good	77.0	56.3	61.9	62.2	78.3	27.3	62.3	29.2	29.5	28.3	30.3	49.3
	Missing	0.4	0.0	1.2	0.0	0.4	3.4	0.2	2.8	1.7	3.8	3.6	1.7
First Language	Afrikaans	1.5	9.2	7.0	8.2	1.6	56.1	9.6	55.2	59.0	58.5	59.2	29.6
	English	7.4	31.3	40.2	40.8	0.7	4.8	32.2	3.9	3.8	3.7	4.0	15.7
	IsiNdebele	2.9	1.5	1.4	1.1	21.8	1.1	1.2	1.6	1.0	1.2	0.0	3.2
	IsiXhosa	14.7	5.3	7.0	5.9	27.8	1.9	4.7	3.4	1.7	2.5	2.3	7.0
	IsiZulu	28.2	18.7	13.7	14.6	6.6	7.6	19.9	8.3	5.2	7.2	7.8	12.5
	Sepedi	0.7	6.0	4.8	5.5	8.4	11.6	5.8	8.1	12.1	9.2	8.0	7.3
	Sesotho	20.8	5.6	4.3	3.6	3.8	6.1	4.5	6.9	7.7	7.2	6.5	7.0
	Setswana	4.0	9.9	7.2	7.3	14.9	1.1	10.1	4.3	4.8	4.5	4.2	6.6
	Siswati	7.7	5.4	1.9	2.1	3.9	1.5	5.1	1.8	1.7	0.8	1.5	3.0
	Tshivenda	7.5	2.4	5.1	5.2	7.2	2.9	2.3	2.6	1.0	1.6	0.8	3.5
	Xitsonga	3.9	3.7	3.1	3.2	3.0	2.9	3.4	2.2	1.0	2.3	3.6	2.9
	Other	0.3	1.0	3.1	2.5	0.3	0.6	1.2	0.4	0.2	0.6	0.4	1.0
Race	Missing	0.4	0.0	1.2	0.0	0.0	1.7	0.0	1.3	0.8	0.7	1.7	0.7
	African	96.7	61.6	55.2	54.4	97.5	37.4	61.0	40.0	37.7	37.8	35.5	55.9
	Indian	0.2	5.4	8.0	7.7	0.2	1.1	5.9	0.4	0.6	0.6	0.2	2.8
	Mixed-race	1.6	5.4	8.0	6.4	1.5	4.0	5.9	3.4	2.5	4.3	3.8	4.3
	White	0.7	27.4	27.5	31.2	0.7	54.8	27.0	53.5	56.3	55.8	57.1	35.6
	Missing	0.8	0.2	1.3	0.3	0.1	2.7	0.2	2.7	2.9	1.5	3.4	1.5

CN, conscientiousness; ES, emotional stability; EX, extraversion; FC, facilitating; IT, integrity; IL, intellect; OP, openness; RH 1, relationship harmony version 1; RH 2, relationship harmony version 2; SH 1, soft-heartedness; SH 2, soft-heartedness version 2.

Firstly, the dataset included more than 50 000 statements in 11 official languages. The authors had to evaluate them and turn them into items. Some of the personality facets they identified were common to all 11 official languages, whilst others were specific to some languages or to only one language. The challenge was to determine which responses were relevant for developing items. Some items seemed to be culture-specific and could have different connotations in different cultural groups. Examples are: 'bewitching people' (soft-heartedness) and 'wandering in the streets' (conscientiousness). Some groups could regard items of this nature as offensive or confusing.

Secondly, some items were too vague or abstract. The authors eliminated them in the process of generating items or revised them by contextualising the specific items. For example, an item pertaining to being 'outgoing' became 'go out to parties' (extraversion). During this phase, the authors eliminated items if they felt that the items did not measure a certain facet or if the context was vague.

Thirdly, the items had to exclude idiomatical expressions, like 'to jump the gun' (extraversion) or 'standing behind friends' (soft-heartedness). Including these items would have been problematic because previous South African studies, which focused specifically on the fair cross-cultural application of instruments, suggested that items should be free from aspects that different groups could understand in different ways (Meiring *et al.*, 2005; Taylor & De Bruin, 2005).

As a result, the authors developed and retained 2573 items to measure the nine clusters. They prepared and administered these items in 11 different studies via paper-and-pencil instruments or internet-based surveys.

The authors measured each of the constructs using separate questionnaires, except for the relationship harmony and soft-heartedness constructs. Because the large number of items they generated for the relationship harmony and soft-heartedness constructs (400 and 482 respectively), the authors developed two questionnaire versions for these clusters to avoid using very long questionnaires (they labelled these questionnaires RH-1 and RH-2 for relationship harmony and SH-1 and SH-2 for soft-heartedness). They identified anchor items and included them in both versions of the questionnaires, whilst they divided the remaining items randomly between the two versions. They rated items on a five-point Likert-type scale, where responses ranged from 1 (strongly disagree) to 5 (strongly agree).

Research procedure

The authors obtained permission to administer the questionnaires and ethical clearance from the different universities and training institutions. They explained the purpose of the SAPI project to the participants and they gave their informed consent by completing and returning a consent form. The authors adhered to all the necessary ethical guidelines when they collected the data through the

surveys and ensured the participants that they would handle the data confidentially.

The SAPI project members collected the data. The authors administered the conscientiousness, integrity, intellect, relationship harmony and soft-heartedness questionnaires using a paper-and-pencil format, whilst they administered the emotional stability, extraversion, facilitating and openness clusters via the internet. Each participant had to complete only one of the 11 questionnaires. Data collection stretched over a period of eight months. Student participants received course credit for participating in the study.

Statistical analysis

The authors performed the statistical analysis using the SPSS package (SPSS Inc., 2010). They inspected the data for missing and/or unexpected values. They checked the minimum and maximum values, as well as the means and standard deviations to determine their plausibility. The authors then investigated the skewness and the kurtosis coefficients of the items from the questionnaires and identified items with an absolute value for skewness of > 2 or for kurtosis of > 4 . They excluded these items from further analyses.

The authors then investigated the item loadings with total score (cluster score) and performed a principal component analysis of items. They requested one component and inspected the component matrix to identify items with absolute loadings of < 0.20 . Although they could have set criteria that are more stringent for the component matrix, they decided to be over-inclusive at this stage of the analyses and to remove only the weakest items systematically.

The authors examined the item loadings with facets and repeated the procedure to determine the item loadings with the total score for the items within the facets of the questionnaires. During the analyses, they selected only the items that they intended to represent a particular facet for the principal component analysis. The authors retained one component and inspected the loadings of the items. The authors expected that all the items would have relatively large loadings (> 0.30). They removed items with low loadings (< 0.30) because they gave early indications that the items were not functioning in accordance with expectations.

Finally, the authors calculated Cronbach's alpha coefficients for the various facets in order to assess the reliability of the facets they measured. According to Frisbie (1988), highly acceptable reliabilities for standardised tests will yield test scores between 0.85 and 0.95. The guidelines that Cicchetti (1994) described suggest that one should regard clinical significance as unacceptable when a reliability coefficient is below 0.70. One should regard reliability coefficients of between 0.70 and 0.79 as fair, reliability coefficients of between 0.80 and 0.89 as good and reliability coefficients of 0.90 or above as excellent. However, one could regard a reliability coefficient of 0.70 or higher as acceptable during the research and when developing instruments (see Nunnally & Bernstein, 1994).

Results

Psychometric properties of the South African Personality Inventory items

The results show that, of the 2573 items the authors developed, 36 items from the Conscientiousness, Integrity, Openness, Relationship Harmony and Soft-heartedness clusters yielded an absolute value for skewness of > 2 and for kurtosis of > 4 . Taking into account the item correlations with the total scores of the various clusters, the authors eliminated 219 items because they shared less than 5% of their variance with the total score. After eliminating these items, only 17 items appeared not to represent the particular facet for which the authors wrote them. The authors excluded these items from further analyses. Finally, the authors also excluded 33 items (including items from all the clusters except Openness), which decreased the reliability of the various clusters, from further analyses. The authors eliminated 305 items during this process (see Table 4).

Reliability

The results showed that the Cronbach's alpha coefficient values of 16 of the facets were lower than 0.70. These facets consisted of between one and four items and were mostly from the soft-heartedness and conscientiousness clusters. The authors excluded these facets from further analyses. The remaining reliability coefficients for the various clusters

ranged between 0.71 (conscientiousness: rebellious) and 0.97 (openness: religiosity).

Factor analysis

The section below presents the various factor solutions for each of the SAPI clusters. Factor analysis determined the construct validity of the clusters and facets.

Firstly, the authors performed a simple principal component analysis on the facets in the various clusters to determine the number of factors to extract. They did this by investigating the clusters' eigenvalues, scree plots and parallel analysis outcomes. Before performing the principal component analyses, the authors assessed the suitability of the data for factor analysis by inspecting the Kaiser-Meyer-Olkin values and the Bartlett's Test of Sphericity.

Table 5 summarises the findings for the nine clusters.

The Kaiser-Meyer-Olkin values for all the scales were acceptable and the Bartlett's Test of Sphericity was significant in all instances. This confirms that the authors could perform exploratory factor analyses on the various data sets.

Depending on the suggestions from the three indicators (eigenvalues, scree plot and parallel analysis), the authors examined one, two, three, four, five or six factor solutions

TABLE 4: Psychometric properties of the South African Personality Inventory items.

Cluster	Number of items developed	Items needed to be reverse scored	Items with extreme skewness or kurtosis	Items that did not load with the total cluster score	Items that did not load with the total facet score	Items that decreased reliability	Remaining items
Conscientiousness	240	12	11	28	0	6	195
Emotional stability	325	0	0	44	6	4	271
Extraversion	379	0	2	37	5	2	333
Facilitating	184	0	0	2	0	1	181
Integrity	117	11	15	12	1	7	82
Intellect	202	0	0	13	0	4	185
Openness	244	0	5	13	0	0	226
Relationship harmony complete ^a	400	4	2	34	3	7	354
Version 1	242 of 400	4	1	20	3	1	217 of 242
Version 2	247 of 400	4	1	16	4	6	220 of 247
Soft-heartedness Complete ^a	482	0	1	36	2	2	441
Version 1	302 of 482	0	1	28	0	0	273 of 482
Version 2	318 of 482	0	0	21	2	2	293 of 482
Total	2573	27	36	219	17	33	2268

^a, Version 1 and Version 2 share these items.

TABLE 5: Factorability of the correlation matrices and the number of factors retained for the South African Personality Inventory clusters.

Cluster	Kaiser-Meyer-Olkin	Bartlett's Test of Sphericity Significance	Eigen-values (Kaiser's criteria)	Scree plot	Parallel analysis	Final factor solution	Percentage of variance explained	α Range
Conscientiousness	0.97	0	2	1	1	1	56	0.92
Emotional stability	0.91	0	5	5	3	4	62	0.77–0.83
Extraversion	0.91	0	5	4	4	3	59	0.79–0.92
Facilitating	0.93	0	2	1	1	1	58	0.93
Integrity	0.91	0	1	1	1	1	58	0.88
Intellect	0.91	0	2	2	2	2	58	0.74–0.84
Openness	0.88	0	4	4	4	1	38	0.86
RH version 1	0.92	0	3	3	3	2	55	0.88–0.91
RH version 2	0.94	0	4	2	2	2	54	0.81–0.92
SH version 1	0.95	0	5	2	3	2	57	0.94
SH version 2	0.95	0	5	4	4	4	66	0.82–0.95

RH, relationship harmony; SH, soft-heartedness.

for the various subscales. Consequently, they used the factor solution that seemed to be theoretically and psychometrically the most sound in each instance.

After determining the number of factors to extract for each of the SAPI clusters, the authors performed maximum likelihood analyses on the range of facets for each of the clusters. The variance that these factors explained ranged between 38% and 66%. The scales the authors derived from these factor-analytic results in each cluster yielded adequate Cronbach's alpha values that ranged from 0.74 to 0.95 (see Table 5).

Discussion

The overall aim of the SAPI project was to develop and validate an indigenous personality instrument that will measure personality in a valid, reliable, equivalent and bias-free way irrespective of peoples' culture, race, language or socioeconomic background (see Nel *et al.*, 2012; Valchev *et al.*, 2011; Valchev *et al.*, 2013).

This study focused specifically on developing an item pool to measure the various personality facets, sub clusters and clusters that the researchers identified in the first stage of the SAPI project (Nel *et al.*, 2012) and on determining which of these items assesses the specific facets, sub clusters and clusters in a realistic and consistent way.

The discussion below focuses on the performance of the 2573 items the authors included in the item, validity and reliability analyses. The reason for this process was to establish which items showed a normal distribution and loaded onto both the facet and cluster the authors intended them to measure. In addition, the analysis aimed to establish the internal consistency between items for each facet. When developing instruments, Worthington and Whittaker (2006) suggest that it is vital to eliminate as many items as possible in the earlier stages of item analysis. It ensures that one measures the factors (clusters in this case) one extracts subsequently using meaningful items for the final instrument.

Psychometric performance of the items

A secondary objective of this study was to establish which items showed a normal or atypical distribution.

According to Rouse (2007), this method of analysis is important because one should identify major disparities in the parameters early in a study in order to disregard biased items before continuing with factor analysis. This may also be the ideal way to identify difficult and discriminative items (Ross, 2005).

The results of this study showed that the authors eliminated only 1.4% of the items because of atypical distribution. The authors identified as many reasons as possible for these results.

For example, the items that related to 'laziness' and 'cleaning' (both from conscientiousness) showed high skewness and kurtosis. The data collection context may have caused these findings.

The authors first collected the data for Conscientiousness at the police services using a sample of entry-level recruits. Most of the participants either agreed or disagreed with both statements in order to show themselves in a favourable light. Therefore, the authors eliminated these items because they may include a socially desirable component that depends on the context of measurement.

Another two types of items, which related to 'proudness of values' and 'proudness of family', also showed atypical distributions (both are from relationship harmony). These items could be problematic because the facet they should have measured was Proud. However, statistical analysis showed that participants might have attributed this pride to two different things (values and family). As a result, the authors eliminated both items.

After this process, the authors conducted a validity analysis on the items (Costello & Osborne, 2005). The objective was to establish whether each item actually measured its intended cluster. Because of this analysis, the authors eliminated 8.51% of the items because they did not correlate with the total score of the cluster. This indicated that certain items did not measure the expected cluster and the authors should have included them in a different cluster or disregarded them.

For example, the authors initially included an item that measures 'to do what you say' in the Conscientiousness cluster. Although this item appeared to measure the facet it should have measured, it failed to measure the overall conscientiousness cluster (within which the facet falls). Therefore, the authors thought that retaining this item was problematic because the objective was to retain items that measure their intended facet and cluster. This item may have fitted better into another cluster, like emotional stability.

When the authors re-assessed the definition of emotional stability, the element of unpredictability seemed to be more relevant for this type of item. However, the emotional stability cluster also included some irrelevant items. Another example of an item that failed to measure the overall cluster was 'to wash hands repeatedly'. Although this type of item (which also assesses compulsive behaviour) seemed relevant for inclusion in the overall emotional stability cluster, it failed to load. This may be because the item assessed a specific element of psychopathology and participants who display that type of behaviour may not be prepared to recognise that element of their behaviour or may be in denial (Andersson *et al.*, 2011; Mataix-Cols, Marks, Greist, Kobak, & Baer, 2002).

After this process, the authors inspected the item loadings to identify items that actually loaded on the specific facet they should have measured. The results showed that only

0.66% of the items failed to load. This showed that these items measured the cluster but not the specific factor they should have measured. Examples from the relationship harmony cluster included items that measured 'strong will' and 'to keep from getting involved in disagreements'. It is possible that the latter item shows social desirability and this could explain the finding that the item did not load onto the intended facet. The first item may be ambiguous as one could evaluate the word 'strong' physically, although this was not its intention. The authors deleted both items from the item pool.

The next objective was to establish the internal consistency of items that measure a specific facet.

The authors eliminated only 33 (1.28%) of the items during this stage. Most facets showed acceptable internal consistency (more than the guideline of $\alpha \geq 0.70$ that Nunnally & Bernstein, 1994) set. The authors either eliminated the facets that showed a lower internal consistency completely or removed the items in these facets that contributed to the lower reliability. Examples of facets that they disregarded during this stage include: gullible and serious (both from soft-heartedness); fair and discriminative (integrity); performance oriented and thorough (conscientiousness); and tolerant (relationship harmony). These facets contained too few items to yield an adequate reliability (like agreeable from soft-heartedness) or their items showed low internal consistency. Therefore, the authors considered these facets as not appropriate enough to measure the specific cluster.

The role of the separate quantitative studies (each of which measured a cluster of the SAPI) was to establish dimensions and their items (Wittes & Brittain, 1990). Of the 2573 items the authors initially developed, they retained 2268 items after the quantitative studies. It meant that they eliminated 305 items and retained 88% of them.

This created a challenge for constructing the SAPI because the authors retained so many items. In the diverse context of South Africa, constructing items is an extremely sensitive undertaking because psychological instruments must adhere to the guidelines that the *Employment Equity Act* (1998) has set. This means that items must measure the same facet in the same way in the different groups and be free of ambiguity, misperception and complexity (Ross, 2005).

This study identified a small number of items that the authors disregarded because they did not meet these criteria. It is also worth noting that most facets showed acceptable internal consistency. This shows that there is low measurement error and that the items measure the intended facet fairly well (Suen & McClellan, 2003).

Additional analyses examined the extent to which the (quantitative) factor analytic results converged with the (qualitative) content analysis of the items, as Nel *et al.* (2012) identified in the previous phase of the project. According to Nel *et al.* (2012), Conscientiousness consisted of five sub clusters.

However, the facets distributed themselves around a single overall factor (or sub cluster). This was also the case with Integrity (originally two sub clusters), Openness (originally four sub clusters) and Facilitating (originally two sub clusters). It seems that these facets interrelate more strongly than was originally perceived. It is also possible that the items that measure the facets are too similar and are, therefore, unable to distinguish between the facets (Norris & Lecavalier, 2010).

Emotional stability included six sub clusters in the qualitative analysis. However, the quantitative analysis yielded only four factors (sub clusters). The four factors are neuroticism, dignified, ill-temperedness and emotional stability. Their current arrangement is similar to the original composition.

The authors retained the Neuroticism sub cluster with a similar formation to the one they identified in the qualitative stage. The other facets distributed beneath the three other factors. Some of the elements that appeared to relate to emotional control, emotional compassion and self-esteem loaded onto the dignified factor (some of these facets initially clustered beneath the emotional control, emotional sensitivity and ego-strength sub clusters (see Nel *et al.*, 2012).

The temperamental facets (or states) loaded onto the Ill-temperedness factor and the more trait-like facets onto the emotional stability factor. These facets were initially composites of the sub clusters of balance, courage, ego-strength and emotional control. Most of the facets that comprised the new sub cluster (labelled emotional stability as well) come from four different sub clusters founded in the qualitative stage. This made the label appropriate.

Extraversion and Intellect each consisted of four sub clusters. However, the authors extracted only three factors (sub clusters) for extraversion and two for Intellect. For extraversion, the factors that the authors extracted were gregariousness, introvertedness and candidness. The extroverted and introverted qualities (initially clustered collectively beneath the Sociability, positive emotionality and dominance sub clusters) appear to have split to form the gregariousness and introvertedness factors respectively. Only the Candidness factor retained the predominant facets of the initial Expressiveness sub cluster. For Intellect, the authors labelled the two sub clusters they extracted Aesthetics and Intellect. All the facets relevant to artistic and creativity appear to have loaded onto the Aesthetics sub cluster, whilst the remaining facets loaded onto the Intellect sub cluster.

In both versions of Relationship Harmony, the authors extracted only two factors, or sub clusters, (Positive Relational Behaviour and Negative Relational Behaviour), although they extracted four sub clusters from the original data (Nel *et al.*, 2012).

For the final cluster, the authors extracted soft-heartedness, two sub clusters from version 1 (detrimental behaviour and good-naturedness) and four sub clusters from version 2 (detrimental behaviour, active support, egoism and amiability).

This differed from the original composition that included six sub clusters. An examination of the differences between the original relationship harmony and soft-heartedness clusters and the current composition of these clusters suggests that the positive and negative facets clustered together in both clusters, thus leading to the extraction of fewer factors.

Therefore, the authors conclude that the overall personality structure that Nel *et al.* (2012) proposed differs to some extent from the personality structure this study found.

It is possible to suggest some reasons for these differences.

The first reason may relate to the nature of the facets that the authors measured. The authors constructed the items so that they asked the participants to evaluate themselves on a 5-point scale by agreeing or disagreeing with statements. During the factor analysis it was evident, on the facet level, that the negative facets grouped together and the positive facets grouped together (specifically in relation to the relationship harmony and soft-heartedness clusters).

A study by Pettersson and Turkheimer (2010) also reported this tendency. They found that, when one includes evaluative terms in an overall personality instrument and the number of negative and positive statements does not balance, the negative facets and positive facets tend to split when one performs exploratory factor analysis. Therefore, Smith (2003) suggested that an individual facet should include an equal number of negative and positive items in order to derive data that are more substantial.

The second reason may relate to including various social desirability elements in the personality structure that Nel *et al.* (2012) identified. In both the relationship harmony and soft-heartedness socially desirable elements (caring, loving and peacekeeping), and socially undesirable elements (abusive, provoking and argumentative) are present. Integrity includes more desirable elements (loyal, truthful and trustworthy) than it includes undesirable ones (like pretending). It is possible that participants were inclined to agree more with the desirable items than with the undesirable ones and this might have distorted the factor analysis of the clusters and sub clusters. Previous research findings indicate that the effects of social desirability may alter the accuracy of a derived personality structure (Bourgeois, Loss, Meyers & LeUnes, 2003).

A third reason may be that some of the facets did not cluster correctly beneath the relevant clusters or showed double loadings onto more than one sub cluster. Three of the facets, particularly musical (intellect), religiosity (openness) and materialistic (openness), failed to load onto the cluster they were supposed to measure. This implies that these facets may be better suited for inclusion in another cluster. Facets that showed double loadings were more prevalent than were those with no loadings. This suggests that the composition of some facets may be uncertain and that one needs to refine them further in order to distinguish these facets. Most of the double loadings centred towards the emotional stability

cluster. The facets short-tempered, depressive and needy loaded onto more than one sub cluster. These facets related to being easily angered, being sad most of the time and needing to be wanted or accepted (Nel *et al.*, 2012).

The development of items for these facets should focus on content distinction to avoid overlap across facets.

Limitations of the study

This study has limitations. The first is that the authors conducted 11 different studies to assess the nine clusters. Therefore, one can make only limited inferences about the overall personality structure, internal validity of the structure and the composition of the facets, sub clusters and clusters in this structure. The 2573 items the authors developed to measure the clusters in the SAPI made a single study unrealistic and served as the motivation for conducting 11 different studies.

Another limitation is the different modes the authors used in the 11 studies (which included using 11 different questionnaires). Participants completed some questionnaires using the paper-and-pencil format, whilst they completed other questionnaires via the Internet. The different modes may have yielded different approaches from participants, depending on the mode with which the particular participant felt more comfortable. According to Davies, Foxcroft, Griessel and Tredoux (2005), computer anxiety may distort results if participants are not computer literate. Using cross-sectional designs in the different studies limited the ability to make interpretations about the causal effect of the items' performance. The authors could have avoided this problem if they had used a longitudinal design.

The sampling method the authors used in most of the studies is another limitation. Using non-probability convenience sampling limited the inclusion of diverse groups in the studies. This further restricted the inferences the authors made about the performance of the items.

Furthermore, this study used only students, limiting the possibility to draw conclusions about maturational effects. However, changes in personality generally refer to mean level shifts rather than to structural changes (see Allemand, Zimprich, & Hendriks, 2008). Therefore, one can expect that the outcomes of the authors' analyses of scale properties will generalise to other age groups.

The final restriction is the use of English throughout the studies. Only 15.7% of the 6735 participants indicated that they used English as their home language. This means that participants may have misinterpreted items, despite the fact that the authors used certain procedures to reduce misinterpretation. These procedures included using cultural and language experts to evaluate items beforehand to determine whether participants from the 11 official languages were able to understand each English statement so that the authors did not include ambiguous and/or dubious items in the questionnaires.

Recommendations

The main recommendation for future research into the SAPI is to include participants who adequately represent different South African groups with regard to their racial, linguistic and cultural diversity. This will assist future researchers to exclude items that show bias and misconception and to establish construct and item equivalence. In order to reduce the number of items further, the authors recommend that language experts translate the remaining 2268 items into all 11 official languages in order to identify items that are not transferable. Researchers can investigate the remaining translated items further in cross-cultural studies in order to identify equivalent and unbiased items. Furthermore, researchers can use the Rasch model for analysing the remaining items. Using a Rasch analysis will allow researchers to extract evidence that is more substantial to demonstrate whether the items fit the facet and the cluster and to separate participants who may have different standings on the facet of interest.

A more practical recommendation for future research is to assess the predictive validity of the SAPI compared to other personality measures in organisational settings. In this study, the authors only determined construct validity. Because the main intention of the *Employment Equity Act* (1998) is to ensure valid, reliable and fair psychological treatment in organisational contexts, a study that compares the SAPI with measures like the occupational personality questionnaire, the 16 personality factors (PF), the 15 Factor Questionnaire (FQ) and other measures is necessary. These measures have been adapted and considerably improved for the South African context over the last 10 years and it is important to assess the incremental value of the SAPI compared to these measures.

Conclusion

This study established that the process of developing, refining and analysing items was more challenging than the authors initially assumed it would be, especially in the South African context. It presented findings that are meaningful with regard to overall test construction and test use in South Africa.

The authors developed items for the SAPI that were close to the descriptive terms the first phase of the project identified (see Nel *et al.*, 2012; Valchev *et al.*, 2013), thus signifying the inclusion of diverse indigenous South African perspectives during item construction.

According to Foxcroft (2009) and Meiring *et al.* (2005), one can prevent item bias if one monitors the setting (or context) when developing items. This may account for the fact that the authors retained 88% of the items after completing the studies.

Apart from the satisfactory performance of the items in the thirteen separate studies, this study also aimed to establish whether one could retain the personality structure (consisting of nine clusters, 37 sub clusters and 188 facets) the first stage of the study proposed (Nel *et al.*, 2012; Valchev *et al.*, 2013) in the quantitative part of the project. The individual studies

showed that some of the clusters did not replicate the cluster as Nel *et al.* (2012) originally proposed. Some facets appear to resemble constructs different to those in the original proposal.

The authors also reduced sub clusters. The original structure consisted of nine clusters and 37 sub clusters whilst the individual studies identified either 17 or 19 sub clusters (depending on which versions of the Relationship Harmony and Soft-heartedness questionnaires the authors used). This suggests that one could reorganise or simplify the proposed personality structure (Nel *et al.*, 2012) in future.

Acknowledgements

The authors acknowledge the financial assistance of the National Research Foundation (NRF; SUR2009062300001496) and SANPAD (09_42) for this research. The opinions the authors express and the conclusions they reach in this article are theirs and one should not attribute them to the National Research Foundation or to the South Africa Netherlands Research Programme on Alternatives in Development.

In addition, the authors want to thank the following people for helping to gather the data: Elke Chrystal, Karina de Bruin, Angelique Flattery, Charnelle Janse van Rensburg, Antoinette Labuschagne-Bierman, Megan Lötter, Talitha Oosthuizen and Petrus van der Linde.

Competing interests

The authors declare that they have no financial or personal relationship(s) that may have inappropriately influenced them when they wrote this article.

Authors' contributions

C.H. (University of Johannesburg) wrote the literature review, research approach, research method and results. She assisted with the data collection as well as the statistical analysis. She reviewed and prepared the final manuscript and was responsible for submitting it. J.A.N. (North-West University) wrote the introduction, discussion, limitations and recommendations in addition to helping collect the data, assisted with the statistical analysis and was responsible for integrating the comments of co-authors and for the technical standard of the final manuscript. F.J.R.V.D.V. (Tilburg University; North-West University and University of Queensland) reviewed the manuscript, gave guidance on matters relating to the literature, statistical analysis and the final manuscript. D.M. (University of Pretoria and University of Stellenbosch) was mainly responsible for the SAPI project management in the SAPI item development phase of steps 1, 2, 3 and 4 in addition to helping collect the data at the South African Police Services and was the supervisor for the conscientious and integrity studies. He also reviewed the final manuscript. V.H.V. (Tilburg University) was centrally involved in developing and implementing the item generation procedures, helped to collect the data, assisted with statistical analysis and reviewed the final manuscript. B.G.A. (University of Johannesburg and Tilburg University) was a member of the core team involved in developing and

generating items. He helped to collect the data, assisted with the statistical analysis and reviewed the final manuscript. G.P.D. (University of Johannesburg) was mainly responsible for providing guidance on the statistical analysis.

References

- Allemand, M., Zimprich, D., & Hendriks, A.A.J. (2008). Age differences in five personality domains across the lifespan. *Developmental Psychology*, 44(3), 758–770. <http://dx.doi.org/10.1037/0012-1649.44.3.758>, PMID:18473642
- Andersson, E., Ljótsson, B., Hedman, E., Kalso, V., Paxling, B., Andersson, G., Rück, C. (2011). Internet-based cognitive behaviour therapy for obsessive compulsive disorder: A pilot study. *BMC Psychiatry*, 11, 125. <http://dx.doi.org/10.1186/1471-244X-11-125>, PMID:21812991
- Angleitner, A., Ostendorf, F., & John, O.P. (1990). Towards a taxonomy of personality descriptors in German: A psycho-lexical study. *European Journal of Personality*, 4, 89–118. <http://dx.doi.org/10.1002/per.2410040204>
- Ashton, M.C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11, 150–166. <http://dx.doi.org/10.1177/1088868306294907>, PMID:18453460
- Avdeyeva, T.V., & Church, A.T. (2005). The cross-cultural generalisability of personality types: A Philippine study. *European Journal of Personality*, 19, 475–499. <http://dx.doi.org/10.1002/per.555>
- Bayoglu, B., Unal, O., Elibol, F., Karabulut, E., & Innocenti, M.S. (2013). Turkish Validation of the PICCOLO (Parenting Interactions with Children: Checklist of Observations Linked to Outcomes). *Infant Mental Health Journal*, 34(4), 330–338. <http://dx.doi.org/10.1002/imhj.21393>
- Bourgeois, A.E., Loss, E., Meyers, M.C., & LeUnes, A.D. (2003). The athletic coping skills inventory: Relationship with impression management and self-deception aspects of socially desirable responding. *Psychology of Sport and Exercise*, 4(2), 71–79. [http://dx.doi.org/10.1016/S1469-0292\(01\)00024-3](http://dx.doi.org/10.1016/S1469-0292(01)00024-3)
- Boyle, G.J., Matthews, G., & Sklofske, D.H. (2008). Personality measurement and testing: An overview. In G.J. Boyle, G. Matthews, & D.H. Sklofske (Eds.), *The SAGE handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2) (pp. 55–72). London, UK: SAGE Publications. <http://dx.doi.org/10.4135/9781849200462.n1>, <http://dx.doi.org/10.4135/9781849200479.n1>
- Cheung, F.M., Cheung, S.F., Zhang, J., Leung, K., Leong, F., & Huiyeh, K. (2008). Relevance of openness as a personality dimension in Chinese culture: Aspects of its cultural relevance. *Journal of Cross-Cultural Psychology*, 39, 81–108. <http://dx.doi.org/10.1177/0022022107311968>
- Cheung, F.M., Van de Vijver, F.J.R., & Leong, F.T.L. (2011). Toward a new approach to the study of personality in culture. *American Psychologist*, 66, 593–603. <http://dx.doi.org/10.1037/a0022389>, PMID:21261408
- Chiu, C.-Y., Kim, Y.-H., & Wan, W.W.N. (2008). Personality: Cross-cultural perspectives. In G.J. Boyle, G. Matthews, & D.H. Sklofske (Eds.), *The SAGE handbook of personality theory and assessment: Personality theories and models* (Vol. 1) (pp. 124–144). London, UK: SAGE Publications. <http://dx.doi.org/10.4135/9781849200462.n6>
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290. <http://dx.doi.org/10.1037/1040-3590.6.4.284>
- Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10, 1–9.
- Davies, C., Foxcroft, C., Griessel, L., & Tredoux, N. (2005). Computer-based and internet delivered assessment. In C. Foxcroft, & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context* (2nd edn.) (pp. 153–166). Cape Town, South Africa: Oxford University Press.
- De Raad, B., Sullot, E., & Barelds, D.P.H. (2008). Which of the big five factors are in need of situational specification? *European Journal of Personality*, 22, 269–289. <http://dx.doi.org/10.1002/per.668>
- Di Blas, L. (2005). Personality-relevant attribute-nouns: A taxonomic study in the Italian language. *European Journal of Personality*, 19, 537–557. <http://dx.doi.org/10.1002/per.569>
- Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling*, 7, 3–28. <http://dx.doi.org/10.1177/1471082X0600700102>
- Fitzpatrick, R., Norquist, J.M., Jenkinson, C., Reeves, B.C., Morris, R.W., Murray, D.W., & Gregg, P.J. (2004). A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. *Quality of Life Research*, 13, 331–338. <http://dx.doi.org/10.1023/B:QURE.0000018489.25151.e1>, PMID:15085905
- Foxcroft, C. (2009). Developing a psychological measure. In C. Foxcroft, & G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (3rd edn.) (pp. 65–76). Cape Town, South Africa: Oxford University Press Southern Africa.
- Foxcroft, C., & Roodt, G. (Eds.). (2009). *Introduction to psychological assessment in the South African context* (3rd edn.). Cape Town, South Africa: Oxford University Press.
- Frisbie, D.A. (1988). Reliability of scores from teacher made tests. *Educational Measurement: Issues and Practice*, 7(1), 25–35. <http://dx.doi.org/10.1111/j.1745-3992.1988.tb00422.x>
- Goldberg, L.R., & Somer, O. (2000). The hierarchical structure of common Turkish person-descriptive adjectives. *European Journal of Personality*, 14, 497–531. [http://dx.doi.org/10.1002/1099-0984\(200011/12\)14:6%3C497::AID-PER379%3E3.0.CO;2-R](http://dx.doi.org/10.1002/1099-0984(200011/12)14:6%3C497::AID-PER379%3E3.0.CO;2-R), [http://dx.doi.org/10.1002/1099-0984\(200011/12\)14:6%3C497::AID-PER379%3E3.3.CO;2-I](http://dx.doi.org/10.1002/1099-0984(200011/12)14:6%3C497::AID-PER379%3E3.3.CO;2-I)
- Government Gazette, 1998, *Republic of South Africa*, Vol. 400. no. 19379, Cape Town, 19 October 1998.
- Hendriks, A.A., Hofstee, W. K., & De Raad, B. (1999). The Five Factor Personality Inventory (FFPI). *Personality and Individual Differences*, 27, 307–325. [http://dx.doi.org/10.1016/S0191-8869\(98\)00245-1](http://dx.doi.org/10.1016/S0191-8869(98)00245-1)
- Ho, D.Y.F. (1998). Indigenous psychologies: Asian perspectives. *Journal of Cross-Cultural Psychology*, 29, 88–103. <http://dx.doi.org/10.1177/0022022198291005>
- Ho, D.Y.F., Peng, S., Lai, A.C., & Chan, S.F. (2001). Indigenization and beyond: Methodological relationalism in the study of personality across cultural traditions. *Journal of Personality*, 69, 925–953. <http://dx.doi.org/10.1111/1467-6494.696170>, PMID:11767824
- Hogan, T.P. (2007). *Psychological testing: A practical introduction* (2nd edn.). Hoboken, NJ: John Wiley & Sons.
- John, O.P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L.A. Pervin & O.P. John (Eds.), *Handbook of Personality: Theory and Research* (pp. 102–138). New York: Guilford Press.
- Mataix-Cols, D., Marks, I. M., Greist, J.H., Kobak, K.A., & Baer, L. (2002). Obsessive-compulsive symptom dimensions as predictors of compliance with and response to behaviour therapy: Results from a controlled trial. *Psychother Psychosom*, 71, 255–262. <http://dx.doi.org/10.1159/000064812>
- Meiring, D., Van De Vijver, F.J.R., & Rothmann, S. (2006). Bias in the adapted version of the 15 FQ+ questionnaire in South Africa. *South African Journal for Psychology*, 36, 340–356. <http://dx.doi.org/10.1177/008124630603600208>
- Meiring, D., Van de Vijver, F.J.R., Rothmann, S., & Barrick, M.R. (2005). Construct, item, and method bias of cognitive and personality measures in South Africa. *South African Journal of Industrial Psychology*, 31, 1–8.
- Nel, J. A., Valchev, V.H., Rothmann, S., Van de Vijver, F.J.R., Meiring, D., & De Bruin, G.P. (2012). Exploring the personality structure in the 11 languages of South Africa. *Journal of Personality*, 80, 915–948. <http://dx.doi.org/10.1111/j.1467-6494.2011.00751.x>
- Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *J Autism Dev Disord*, 40, 8–20. <http://dx.doi.org/10.1007/s10803-009-0816-2>
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. (3rd edn.). New York, NY: McGraw-Hill.
- Pettersson, E., & Turkheimer, E. (2010). Item selection, evaluation, and simple structure in personality. *Journal of Research in Personality*, 44, 407–420. <http://dx.doi.org/10.1016/j.jrp.2010.03.002>
- Ross, D. (2005). *Economic theory and cognitive science: Micro explanation*. Cambridge, MA: MIT Press.
- Rouse, J. (2007). *Practice Theory*. Division I Faculty Publications. Paper 43. <http://wescholar.wesleyan.edu/div1facpubs/43>
- Saucier, G. (2008). Measures of the personality factors found recurrently in human lexicons. In G.J. Boyle, G. Matthews, & D.H. Sklofske (Eds.), *The SAGE handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2) (pp. 29–54). London, UK: SAGE Publications.
- Shaughnessy, J.J., & Zechmeister, E.B. (1997). *Research methods in psychology* (4th edn.). New York, NY: McGraw-Hill.
- Smith, T.W. (2003). Developing comparable questions in cross-national surveys. In J. A. Harkness, F.J.R. Van de Vijver, & P.P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 69–92). New York, NY: John Wiley & Sons.
- SPSS 18.0 [Computer software] (2010). Chicago, IL: SPSS Incorporated.
- Struwig, F. W., & Stead, G.B. (2001). *Planning, designing and reporting research*. Cape Town, South Africa: Maskew Miller Longman.
- Suen, H.K., & McClellan, S. (2003). Item construction principles and techniques. In N. Huang (Ed.), *Encyclopedia of vocational and technological education* (Vol. 1) (pp. 777–798). Taipei, Republic of China (Taiwan): ROC Ministry of Education.
- Taylor, I.A. (2000) *The construct comparability of the NEO-PI-R questionnaire for black and white employees* Unpublished doctoral thesis, University of the Free State, Bloemfontein, South Africa.
- Taylor, N., & De Bruin, G.P. (2005). *Basic Traits Inventory: Technical manual*. Johannesburg, South Africa: J.P. van Rooyen & Partners.
- Valchev, V.H., Nel, J.A., Van de Vijver, F.J.R., Meiring, D., De Bruin, G.P., & Rothmann, S. (2013). Similarities and differences in implicit personality concepts across Ethnolinguistic groups in South Africa. *Journal of Cross-Cultural Psychology*. <http://dx.doi.org/10.1177/0022022112443856>
- Valchev, V.H., Van de Vijver, F.J.R., Nel, J.A., Rothmann, S., Meiring, D., & De Bruin, G.P. (2011). Implicit personality conceptions of the Nguni cultural-linguistic groups of South Africa. *Cross-cultural Research*, 45(3), 235–266. <http://dx.doi.org/10.1177/1069397111402462>
- Van de Vijver, F.J.R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology*, 31, 33–51.
- Visser, D., & Viviers, R. (2010). Construct equivalence of the OPQ32n for Black and White people in South Africa. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 36(1), Art. #748, 11 pages.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wittes, J., & Brittain, E. (1990). The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine*, 9, 65–72. <http://dx.doi.org/10.1002/sim.4780090113>
- Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34, 806–838. <http://dx.doi.org/10.1177/0011000006288127>