EXAMINING AND REVISING THE VOCATIONAL RATING SCALE USING PRINCIPAL COMPONENT ANALYSIS AND MOKKEN SCALE ANALYSIS

REINOUD TIEL GROENESTEGE

University of Twente

Masterthese

Author	Reinoud Tiel Groenestege
Student number	s0151629
Study	Master Psychology
Specialization	Organizational psychology
Organization	University of Twenthe
Supervisors	D. Stoel & Prof.dr. K. Sanders
Date	17th April 2012

CONTENTS

EXAMINING AND REVISING THE VOCATIONAL RATING SCALE USING PRINCIPAL COMPONENT ANALYSIS AND MOKKEN SCALE ANALYSIS

REINOUD TIEL GROENESTEGE

University of Twente

This study attempts to clear up the smoke surrounding the dimensionality of The Vocational Rating Scale (VRS; Barret & Tinsley 1977). As a general measure of self-concept crystallization it has been proven useful in the field of career development. The VRS was developed to measure three of Super's (Super, Starishevsky, Martlin & Jordaan, 1963) metadimensions of the self-concept: clarity, certainty and structure. Little is known however about the underlying factor structure. The current study used principal component analysis (PCA) and Mokken scale analysis (MSA; Mokken, 1971) to test whether the VRS contained three separate factors or was functioning as a one-factor model. The PCA and confirmatory MSA applied to the VRS suggested that the scale was unidimensional. Further analysis using an adaption of the standard exploratory MSA procedure resulted in three separate scales. These scales were validated using MSA and proved to have strong psychometric properties. The scales that were found have implications for the theory of self-concept crystallization.

1. INTRODUCTION

In a dynamic economy and labor market it is of paramount importance for an employee to be self-reliant. A career path nowadays is paved with a variety of jobs rather than one stable employment at one organization. This changing character of the labor market is parallel to a switch from an organizational perspective on career development, to a more individual focus (Baruch, 2004). Organizational structures supported traditional career development, were hierarchy based and offered limited promotional opportunities. Although the organization still plays an important role, the main focus is on the acting individual. This has made researchers come up with career models that focus on the role of the individual in designing its own career. A product of this stream of research is the concept of the *boundaryless career* (Defillippi & Arthur, 1994). A boundaryless career is a career that is not confined by specific organizations and which growth is based on competency. To flourish in a boundaryless career it is hypothesized an individual needs to have three competencies: *knowing why*, *knowing how* and *knowing whom* (Eby, Butts and Lockwood, 2003). A dominance analysis showed that 'knowing why' was the strongest predictor of success in a boundaryless career. According Eby, Butts and Lockwood 'knowing why' on its own term is constituted by three competencies*: proactive personality, openness to experience* and *career insight.* Career insight captures the extent to which a person has knowledge of his or her strengths and weaknesses, has realistic career expectations and clear career goals (Noe, Noe, & Bachhuber, 1990). Of the three 'knowing why' predictors, career insight was best able to predict perceived career success and internal marketability.

A clear lesson learned from the aforementioned research is that not every individual is equality equipped to succeed in a dynamic and apparently boundaryless system. The odds of career success vary with the insight into one's career. A clear picture about who you are, what you want, what your abilities are and in what job you see yourself, is positively associated with career success. It was Super, Starishevsky, Martlin and Jordaan (1963) that in the early 60's already proposed that having a clear sense of self would help choosing the right career and coined the term *vocational self-concept*. Barret and Tinsley (1977a) utilized part of this theory and developed the Vocational Rating Scale (VRS). The VRS was developed to measure three of Super's (Super et al., 1963) metadimensions of the self-concept: *clarity, certainty* and *structure*. Clarity refers to the degree of sharpness of a person's view on his work related attributes. Certainty can be regarded as the degree in which a person is certain about his work related traits. Structure is defined as the differentiation among the self-concepts.

Vocational self-concept crystallization has been shown to relate to important career behavior. In one study by Tokar, Withrow, Hall and Moradi (2003), vocational self-concept crystallization was found to be negatively related to career indecision. In another study, Taylor (1985) found that a more crystallized vocational self-concept predicted a higher satisfaction with the type of work students chose. Lunnenborg (1978) concluded that vocational self-concept crystallization was positively related to the planning style of career decision-making. The planning style involves taking responsibility for one's decision making and doing this in a rational way. Furthermore this style involves introspection and realistic self-appraisal (Lunnenborg, 1978). Recently Weng and McElroy (2010) found that a higher score on the VRS resulted in higher job decision effectiveness. Vocational self-concept crystallization also mediated the effect of career self-management on job decision effectiveness. These studies indicate that vocational self-concept crystallization is a useful construct in the field of career development. However, the VRS is subject to some concerns.

First of all, all studies report a very high alpha of .94. A high alpha is usually presented as a merit, but an alpha as high as this can also indicate that some of the items measure the exact same thing and thereby making these items superfluous.

Another issue is the lack of knowledge about the underlying factor structure of the VRS. Tinsley, Bowman and York (1989) conducted a factor analysis on four vocational choice questionnaires: Career Decision Scale, My Vocational Situation, Vocational Rating Scale and the Decisional Rating Scale. The fact that four questionaires were included for analysis makes it hard to draw conclusions about the dimensionality of the VRS as a separate scale. Furthermore, the authors used principal component analysis (PCA). Although frequently used to study dimensionality, this method is subject to several limitations that will be further explained in paragraph 2.2.2.

It is not unlikely that the VRS is actually measuring one single construct. Although clearly separate as theoretical constructs, *clarity* and *certainty* might be hard to separate psychometrically. Taking the above into consideration there is much to gain when further studying the VRS. This gain is not limited to a mere theoretical level. If separate constructs like clarity and certainty are indeed psychometrically distinguishable, it becomes possible to train employees or students on those specific dimensions.

Given the ambiguity of the dimensionality of the scale we requested to reanalyze the original data of Tinsley, Bowman and York (1889). With the original data it's possible to make clean comparisons between results. When it comes to dimensionality research, PCA is the most popular method. Because PCA suffers from several limitations, this study will also use a method derived from Item Response Theory: Mokken Scale Analysis (MSA; Mokken, 1971). This method is relatively underused in scale development but has often produced fruitful results (Wismeijer, Sijtsma, Van Assen & Vingerhoets, 2008; Emons, Sijtsma & Pedersen, 2010). Wismeijer et al. (2008) stated that PCA and MSA can work in a complementary fashion.

The study of dimensionality has always been surrounded by the debate on which method would be best for the job. This debate is nurtured by the fact that each method produces different results. For this reason the current study will utilize both PCA and MSA for examining the dimensionality of the VRS. The main aim of this study is to uncover the true dimensionality of the VRS using both methods, making it able to draw strong conclusions. In doing so, this study will provide a detailed introduction into MSA and introduce a newly devised procedure for executing this particular analysis. It is important to note that this study will not attempt to compare MSA and PCA on a methodological level. Before turning to the analysis of the data, attention will be focused on the origin of the VRS and more light will be shed on the theory of the vocational self-concept and its metadimensions. Next will be a short introduction into PCA followed by a more extensive description of MSA. This distinction in detail was purposefully made due to the fact that MSA is relatively unknown. Before I continue I would like to thank Dr. Tinsley for allowing me to use the original data collected in Tinsley, Bowman and York (1989).

2. THEORY

2.1 Vocational self-concept

"In expressing a vocational preference (Super, 1951), a person puts into occupational terminology his idea of the kind of person he is; that in entering an occupation, he seeks to implement a concept of himself; that in getting established in an occupation he achieves self-actualization. The occupation thus makes possible the playing of a role appropriate to the self-concept." (Super et al., 1963, p. 1)

This quotation starts off the first of five essays written by Super (Super et al., 1963) and illustrates the relevance of this theory and consequently shows the relevance of this study. Presuming self-actualization is the ultimate goal for most people (Maslow, 2006), and at the same time considering the role of the self-concept in reaching this goal, it helps to grasp the importance of inquiry on this subject. Super concluded that the essence of vocational development is comprised of developing and implementing a self-concept. What is a self-concept and, how does it play a role in vocational development?

2.1.1 From' self' to 'vocational self-concept'. Super begins at *the self*. The self is the person and this person can change through interaction with the environment. *Self-percepts* are products of the observations a person makes of himself. These observations are made with the help of the senses. For example, a person can look at himself and see that he is tall. Self-percepts acquire meaning when they are linked to other percepts that not pertain to the self per se. They are linked for example through generalization or comparison. If a self-percept acquires meaning through percepts of other things (other people or believes) the percepts become lower-level or simple concepts. When the link between meaning and percept tightens, the percept becomes a concept. Thus there are two types of self-percepts: primary self-percepts (observations through the senses) and secondary self-percepts (percepts that acquired meaning with the help of other percepts).

A *self-concept* is created when self-percepts holding meaning are related to other self-percepts. Self-concepts can differ in complexity. When a reflection in the mirror is compared with other people's appearances the simple self-concept "tall" comes into being. If you combine this with other simple self-concepts like fast and strong, a person might turn this into a complex concept of a professional basketball player. It is important to keep in mind that not everyone has just one self-concept. Each person has several and these together make up the *self-concept system*. The vocational self-concept is defined as follows: "*The constellation of self attributes which the individual considers vocationally relevant; these may or may not have been translated into a vocational preference."* (Super et al., 1968, p. 19)

2.1.2 Anatomy of the self-concept. With the concept defined, measurement comes into view. When it comes to measuring the self-concept, Super notes that it is of paramount importance that complex self-concepts are not inferred by outsiders. That is to say, complex self-concepts may not be inferred by using the item scores on lower level self-concepts. All that is measured should be the object of awareness. The dimensions of the self-concept that are vocationally relevant are identical to the dimensions of personality. If you want to measure how well a person fits within an occupational role, the focus should be on the personality. Contributing to the study of personality is not what Super intended. Other dimensions of the self namely, the *metadimensions* of the self-concept are the ones that deserve attention and research. The metadimensions are "*characteristics of the traits which people attribute to themselves".* In other words, the characteristics of the self-concepts they acquired. Like the self-concept, the self-concept system also has metadimensions. Super proposed the following thirteen metadimensions:

Self-concepts	*Self-concept systems*
1. Self-esteem	1. Structure
2. Clarity	2. Scope
3. Abstraction	3. Harmony
4. Refinement	4. Flexibility
5. Certainty	5. Idiosyncrasy
6. Stability	6. Regnancy
7. Realism.	

Self-esteem is the appraisal of one's own worth. The amount of self-esteem a person has about a self-concept could be measured with an item like "I am good with people and that makes me happy". *Clarity* of the self-concept: How clear is the person's view of his attributes? Super suggests that having a high degree of clarity might result in having a vocational preference and ease of making a vocational decision. Clarity (which increases with age) is a basic metadimension. A clear self-concept is required before one can measure other metadimensions like *abstraction* for example*.* Abstraction involves the degree to which people are able to abstract from concrete simple self-concepts to higher, more complex self-concepts. The extent to which people are able to elaborate on their traits in a detailed manner is called *refinement. Certainty* refers to the amount of confidence people have in their own traits, the extent to which they are certain of their self-picture. Certainty also plays a big role in vocational choices. One has to be certain of his self-concepts before he will base choices on them. *Stability* denotes the range of consistency of self-concept over time. Having a stable conviction about one's traits may lead to making better decisions. *Realism* of the self-concepts involves the degree of harmony between ones self-concepts and the outside world.

Super suggested six metadimensions for the self-concept system. The abovementioned metadimensions were applicable to any lower or higher level self-concept. The following metadimensions are concerned with the construction of the self-concepts. S*tructure* touches upon the way in which the concepts are internally differentiated. Some concepts can be really descriptive of the self while others don't say much about the person. *Scope* deals with the number and variety of the traits. This is not to be confused with refinement which asks how detailed a person can describe separate traits. Scope only looks at the size of the self-concept collection. If a person is able to adapt his self-concept system adequately when encountering new experiences and new self-percepts, his system is highly *flexible.* Super suggest that flexibility might be a determinant of realism. This sounds plausible, considering that when a person is able to absorb new data into his self-concept in an adequate manner, he will be less likely to lose touch with reality. *Harmony* describes the degree of accord between self-concepts. A harmonious self-concept system is devoid of conflicting concepts. This will make it easier to form a vocational preference. *Idiosyncrasy* might also be called originality. It pertains to the differences between a person's descriptions of himself and that of other people. The last metadimension of the self-concept system was named *regnancy.* It deals with the resilience of a self-concept. Does the self-concept persist even when it conflicts with others?

Super et al. (1963) presents these definitions as "*paving the way"* for the formulation and testing of meaningful hypotheses. Barret and Tinsley (1977a) chose to ride this way and selected three of the abovementioned metadimensions to be operationalized in the Vocational Rating Scale.

2.1.3 Developing the VRS. Fourteen years after Super et al. (1963) first proposed his framework of metadimensions Barret and Tinsley (1977a) note two shortcomings of the self-concept research within the vocational field. First of all: there is hardly any research done on the metadimensions. Furthermore the measures being used to tap into metadimensions like clarity and certainty are subject to two limitations. First, most results are based on externally based inference. That is to say, sum scores translate lower order concepts into higher order concepts. The way lower order concepts form a higher order concept is only known to the person under observation and cannot be inferred by outsiders. Second, the measures only focus directly on lower level concepts and not higher level ones. A problem with Super's framework is that it holds 13 separate metadimensions. Bailey (1970) conducted a research on the independence and factor structure of the metadimensions. He concluded that even though some metadimensions lacked clear independence, each one showed a sufficient amount of individuality to keep regarding them as separate concepts. A principal components factor analysis showed that a simpler structure of five factors could be applied that still would explain 65% of the variance. Apart from the fact that research on the validity of Super's framework produces mixed results, researchers are not able to differentiate the 13 distinctions on an operational level. Barret and Tinsley did not intent to utilize Super's sophisticated framework as a whole but rather develop a scale that could prove the existence of a non-self-esteem metadimension of the self-concept. In line with bailey's conclusion that a simpler model could also be applicable, Barret and Tinsley (1977a) concluded that it would be useful to focus only on three basic metadimensions: clarity, certainty and structure. Together these three would measure the *crystallization of the self-concept*. "*Vocational self-concept crystallization can be defined as the degree to which the constellation of self-attributes which the individuals considers to be vocationally relevant is well formulated. "well Formulated" here refers to the degree to which the separate vocationally relevant self-concepts possess clarity and certainty for the individual, and the constellation of self-concepts as a whole possesses internal differentiation or structure"* (Barren & Tinsley, 1977a, p. 307). This definition is in line with Super's denotation of clarity, certainty and structure. The VRS was designed as a global measure of vocational self-concept crystallization. The scale consists of 40 items that asks a person about his or her vocational attributes and characteristics. All of the items are rated on a 5-point Likert scale that ranges from 'completely false' to 'completely true' (Barret & Tinsley, 1977a). High scores indicate a high degree of vocational self-concept crystallization. Several item examples are displayed below:

I am certain that my knowledge of my own interests, abilities (...) is accurate.

I have a clear idea of my own needs and desires with respect to a career.

I'm not certain about what type of job environment I'd really be happy in.

I know pretty much what I'm looking for in a college major and a career.

The study of Tinsley, Bowman and York (1989) is the only study that examined the dimensionality of the VRS. Analysis of the scales showed three separate factors: *Crystallization, Decision-Making Obstacles,* and *Indecision*. Factor analysis of item responses revealed five factors of which only the first three were stable. The factors were named *clarity, certainty, indecision, decision-making obstacles,* and *informational deficit.* The scale and item structures were related. Results also showed that the VRS was measuring both clarity and certainty, this being in line with the theory. However, this study did not report which items belonged to the clarity construct and which items fell under the certainty construct. This makes it impossible to conclude whether the two constructs are psychometrically separable. Structure did not emerge as a separate construct which is surprising to say the least considering that this is one-third of the final concept. Another important consideration is the fact that Tinsley, Bowman and York (1989) used Principal Component Analysis. Using this method on polytomous items can in many cases cause overfactoring (Fabrigar, Wegener MacCallum ,& Strahan, 1999).

The one thing that is repeatedly stressed in the studies that use the VRS, is its high internal consistency (Cronbach's $\alpha = 0.94$). However, alpha says nothing about the *homogeneity* of the scale. Homogeneity refers to the unidimensionality of a set of items. High internal consistency is a necessary condition for homogeneity but is not sufficient. This argument in combination with the fact that the principal component analysis performed by Tinsley, Bowman and York (1989) did not allow for any definitive conclusions about the VRS's dimensionality directed this study toward the following hypotheses: (a) the VRS contains a single dimension (H0) or (b) upon analyzing the data, the three dimensions (clarity, certainty and structure) proposed by Barret & Tinsley (1977a) become apparent (H1). In the following section effort is devoted to elaborating on the two methods that will be applied to test these hypotheses.

2.2 Dimensionality

2.2.1 Classical test theory. Methods derived from classical test theory (CTT) were the building blocks of test development in the 20th century. According to Embertson & Reise (2000) it all started with the procedures developed by Spearman (1907). CTT is also called *true score theory.* The unit of focus within CTT is the individual *observed score* on a test. The observed score on a test X_i is composed by the accumulation of a person's *true score* T_i and the person's error on the measurement E_i. This results in the following equation:

$$X_i = T_i + E_i$$

The true score can also be called the *trait score.* The trait score indicates the location of a person on the latent variable. With this model there come two assumptions. The first assumption holds that, if a person would complete a test an infinite number of times, the mean error of that test would be zero. This is because errors are considered to be random. The second assumption holds that the person's error is not related to other errors or true scores. These assumptions are regarded to be "weak" because most data sets will satisfy them (de Ayala, 2009). This model has two important limitations. First, the instrument characteristics influence the person's observed score. A very difficult test administered an infinite number of times will produce a different T_i than a very easy test would, also administered an infinite number of times. In this case only the instruments characteristics influence the person's scores. Second, sample characteristics can influence measures of item difficulty and internal reliability. Two samples taking the same test (of let's say 10 items) can produce different alpha's. Furthermore, if one sample would be significantly better at the specific test, then some items might be labeled as 'easy' while based on the results of the other sample the same items might be labeled as 'difficult'. This sample dependence makes the measures of CTT fallible.

2.2.2 Principal Component Analysis One of the techniques to explore the dimensionality of a test is principal component analysis (PCA). PCA sprouted from CTT and is the most commonly used method for inferring the number of factors (from now on interchangeably used with components) underlying a certain dataset. The Kaiser criterion of computing eigenvalues greater than 1.0 is the most popular criterion (Fabrigar et al., 1999). The numbers of eigenvalues greater than 1 constitute the number of factors. An eigenvalue indicates how much additional variance can be explained by the extra factor.

The rationale behind the Kaiser criterion is that a factor should at least explain the equivalent of one original variable. Note that the Kaiser criterion is arbitrary to an extent. It is not really consequential to label an eigenvalue of 1.01 as a "major" factor and totally disregard an eigenvalue of 0.99 (Fabrigar et al., 1999). One problem of PCA is that it can result in *difficulty factors* (Bernstein & Teng 1989; Wismeijer et al., 2008). PCA uses inter-item correlations and this leads to the assumption that relationships between item scores are linear. Categorizing item scores into Likert scales for example creates differences in distribution between item scores. When these differences of distribution exist, the inter-item correlations can no longer be relied upon. If two items are the same in content but differ in distribution, they will share a low correlation (Bernstein & Teng, 1989; Wismeijer et al., 2008). These difficulty factors will in many cases cause *overfactoring* and in some cases *underfactoring* (Fabrigar et al., 1999). See paragraph 2.3.1 for an example of this. It is sometimes suggested to use tetrachoric correlations because these correlations correct for difficulty factors. However, using tetrachoric correlations has its own disadvantages (Embretson & Reise, 2000).

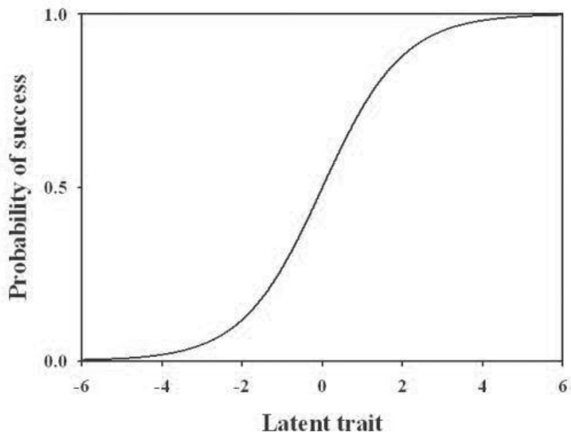
The scree test (Cattell, 1966) is an alternative method for determining the number of factors. Eigenvalues and number of factors are depicted in a scree plot. Every time a successful factor is extracted the eigenvalues become smaller. Once an ideal number of factors are extracted the line is almost horizontal. The researcher looks for the elbow (gap) in the curve to determine the number of components. This approach has a better reputation than the Kaiser criterion but has also been shown to produce ambiguous results when trying to determine the number of factors (Berstein & Teng, 1989; Fabrigar et al., 1999). The biggest problem of the scree test is that it is subjective. The factor analyst is forced to make an 'eyeball' decision (Finch & West, 1997).

The two abovementioned techniques suggest to the researcher how many components should be extracted for *rotation*. As a researcher you want a clear picture of the correlations between the variables and the components. This will make it easier to make theoretical sense of the components. However, before rotation it's hard to interpret a component. Rotation simplifies the solutions so they are easier to interpret. There are two basic types of rotation: orthogonal rotations that do not allow factors to correlate and oblique rotations that do allow factors to correlate (Hatcher, 1994). Although the orthogonal rotation called varimax is most widely used several researchers have argued that using oblique rotation is preferable (Fabrigar et al., 1999; Conway & Huffcutt, 2003). They argue that if factors are indeed correlated, orthogonal rotation creates an unrealistic solution. Ease of interpreting does not guarantee that you make the correct conclusions. Forcing the data into an uncorrelated factor structure does not help to get a clear view of the true dimensionality. Oblique rotation on the other hand will simply show the true data structure. If the factors are in reality uncorrelated both rotation types will be adequate.

2.2.3 Item Response Theory. In the last decennia IRT has become the mainstream theory to turn to when it comes to measurement issues. It has been widely used in education as a means of refining exams, creating item banks for future exams, equate different exam forms and computer adaptive testing (de Ayala, 2009). IRT is sometimes regarded as the stronger versions of true score theory because, as mentioned before, IRT makes stronger assumptions than CTT. Whereas CTT focuses on the observed total score, the unit of focus within IRT is the item of a particular instrument. IRT consists of mathematical models that describe the probability of a response to a particular item. This response is a function of the underlying or latent trait and the characteristics of the item (Chang & Reeve, 2005). Every item has its own item characteristic curve (ICC) in which the probability of getting an item 'wrong' or 'right' is expressed mathematically. For polytomous items the ICC depicts the probability of responses in each category. An example of an ICC for a dichotomous item can be seen in figure 1.

FIGURE 1

ICC of a dichotomous item



As can be seen, the higher the level of the latent trait, the higher the probability of a 'correct' response. IRT generally follows three assumptions (de Ayala, 2009). The first holds that there is a single latent trait which is mathematically expressed by the symbol θ. This is also called *unidimensionality*. Note that multidimensional IRT models do not include this assumption. The second assumption is that of *local independence.* It holds that items are independent from each other. Thus the response of a person to an item is only influenced by the latent trait and not by other items on the test. If this was the case, thus a person's response did get influenced by something other than his or her level of the latent trait, then unidimensionality is no longer the case (de Ayala, 2009). The third assumption is often referred to as *functional form*. It states that the data, if articulated graphically, should follow the form prescribed by the model. For example, figure 1 is an ICC that follows the form of the one-parameter-logistic model (1PL) also known as the Rasch model (Rasch, 1960). This is called the one-parameter-logistic model because it only takes one parameter (item difficulty) into account. The two-parameter-logistic-model (2PL) also includes *item discrimination* in the equation. The three-parameter-logistic-model (3PL) includes a third parameter referred to as pseudo chance (guessing). This parameter addresses the issues of subjects answering correctly by mere chance (Change & Reeve, 2005).

2.2.4 Advantages of IRT. Using IRT models in the area of constructing and testing scales offers several advantages over the traditional CTT methods (Embertson & Reise, 2002; Magno, 2009). Some of these advantages will be named in the following section. For a more detailed discussion see Embertson and Reise (2000). One important advantage of IRT models is the fact that they are neither test nor sample dependent (de Ayala, 2009; Magno, 2009). Another advantage of IRT is that it can predict how a subject is going to respond based on his location on the latent variable. This capacity allows for computerized adaptive testing (de Ayala, 2009). IRT has also shown to be better capable of testing for item bias. In addition, the analyses of the item parameters allow the researcher to order persons on every level of the trait. Estimations concerning the reliability of trait level estimates are made on an individual level rather than based on populations (Rouse, Finger & Butcher, 1999).

2.2.5 Nonparametric IRT. Within the IRT literature an increasing amount of attention is being paid to one particular class called *nonparametric item response theory* (NIRT). Several researchers have argued the importance of studying the uses of NIRT in scale construction and scale evaluation (Meijer & Baneke, 2004; Sijtsma, Emons, Bouwmeester, Nyklíček & Roorda, 2007; Junker & Sijtsma, 2001). NIRT offers several advantages over parametric IRT (PIRT). First, the NIRT models are less restrictive, allowing more items into the scale while at the same time keeping the desirable measurement properties. The ICC's are not required to follow the particular form prescribed by the model. This will prevent researchers from drawing conclusions prematurely, unjustly discarding items and forcing the data into a structure it does not possess (Meijer & Baneke, 2004). Second, NIRT data is easily obtainable by user-friendly software packages. Third, NIRT can also be applied to small samples (Junker & Sijtsma, 2001). Fourth, NIRT helps to understand what PIRT models do (Junker & Sijtsma, 2001). Finally, the NIRT models tend to be relatively simple compared to the PIRT models. Pursuing simpler explanations is preferred in the light of *Occam's razor* that holds that simplicity is a goal itself (Domingos, 1999).

The aforementioned researchers are not advocating that NIRT should replace PIRT in any way. They do however argue that it is better in some cases to use NIRT before applying PIRT methods (Meijer & Baneke, 2004). NIRT is a flexible and data-driven bottom-up approach and can be used even when the data does not fit the prescribed PIRT models. One NIRT model that has proven very useful in the area of scale development and scale revision is the Mokken Scale Analysis (MSA; Mokken, 1971). MSA benefits from all the advantages of NIRT mentioned above. It is an elegant method because its assumptions are based on an underlying model but at the same time it is not subject to the restrictive assumptions that come with most IRT models. In the present study it is the MSA that will be applied for evaluating and possibly revising the VRS.

2.3 Mokken Scale Analysis

2.3.1 From Guttman to Mokken Scaling. Schuur (2003) showed how the traditional measures of reliability can cause a scale to be wrongfully interpreted as multidimensional. Schuur explained this with the following example. A hundred subjects answered to 6 dichotomous items measuring political participation. People were asked if they engaged in certain political activities and could answer either *yes* or *no.* The first three items were rather "unpopular" or difficult. Not many people answered yes to these items. The last three items were rather "popular" or easy. The data was depicted in a table and formed a perfect unidimensional *Guttman scale*. A Guttmann scale is formed when person gives a positive response to a difficult item and has done so as well with all the items which are less difficult. This creates a cumulative scale. See table 1 for an unrelated example of a perfect Guttman scale. In Schuur's example, Cronbach's α of the scale was .84. It could however be increased to .98 if the (difficult) first three items were removed. A standard factor analysis using principal component analysis with VARIMAX rotation also showed two separate factors. So in this case, a scale could be wrongfully judged as multidimensional. This is mainly because traditional methods from CTT assume that data processes normality and that items have the same frequency distribution. In reality however, items don't always have the same frequency distribution. In other words, not all items have the same "popularity". If this is the case, classic measurements of reliability or factor analysis could produce incomplete or misguided results.

TABLE 1
Example of a perfect Guttman scale

Subject	Item				
	1	2	3	4	5
A	0	0	0	0	1
B	0	0	0	1	1
C	0	0	1	1	1
D	0	1	1	1	1
E	1	1	1	1	1

If items are expected to form some sort of a Guttman scale, it seems appropriate to analyze them with the assumption of cumulatively in mind. A perfect Guttman scale does not allow errors, which makes practical use of the idea infeasible. The fundamental idea behind it is highly useful however. There are two approaches to imperfect Guttman data (See table 2 for an imperfect Guttman scale). The first one suggests that deviations from a perfect Guttman scale are systematic and have theoretical value. IRT stood at the base of the second approach of which advocates suggest that the deviations are random and should be treated in a probabilistic manner. The 1PL model is a development that originated from this approach. Although useful, the assumptions of the model were relatively strict and sound application required a high number of items. Mokken scaling was developed to deal with these problems (Schuur, 2003). Mokken scaling offers two models: the *monotone homogeneity model* (MHM) and *the double monotonicity model* (DMM). This study will apply the first model.

TABLE 2
Example of a Guttman scale with Errors

Subject	Item				
	1	2	3	4	5
A	0	0	1	0	1
B	1	0	0	1	1
C	0	0	1	1	1
D	0	1	1	0	1
E	0	1	1	0	1

2.3.2 Assumptions. The monotone homogeneity model (MHM) makes three fundamental assumptions (Sijtsma & Molenaar, 2002).

1. *Unidimensionality;* this means that all items measure the same latent trait denoted as θ. Thus all items in the VRS should measure the extent to which the vocational self-concept is crystallized. Unidimensionality is desirable because it easier to compare examines on the basis of one single score (van Abswoude, van der Ark, & Sijtsma, 2004)
2. *Monotonicity;* that is to say, the ICC (in this context called the item response function (IRF)) is nondecreasing. Thus when the subject's value of θ is increasing, the propability of a positive response to a particular item is at the very least not decreasing.
3. *Local Independence*; means that the responses to items are only influenced by the latent trait and not by answers to previous items.

The double monotonicity model has a fourth assumption. It assumes that the IRF's are nonintersecting. This is translated to subjects having the same interpretation of the item difficulty (Sijtsma & Molenaar, 2002).

Comparison of the assumptions inherent to this nonparametric model with the general assumptions of parametric IRT models shows that the functional form assumption of the PIRT is replaced by the assumption of monotonicity. This reflects the less restrictive character of the MHM model. Instead of requiring the IRF's to follow a specific form, monotonicity only requires them to be nondecreasing. The combination of monotonicity and unidimensionality guarantees that persons can be ordered on a scale based on the latent variable when items are dichotomous. The same assumptions also support this ordering when items are polytomous. This can be accompanied by small unimportant deviations (Emons, Sijtsma, & Pedersen, 2010). Sijtsma and Molenaar (2002) argue that person ordering precision would be higher if all items that have nondecreasing IRF's are included in the scale. Person ordering is important because it allows tests to order people on a scale (Wismeijer et al., 2008). In the present study this means that persons can be ordered based on their value of the latent trait, self-concept crystallization. This can be advantageous for counseling psychologist and assessment centers.

2.4 PCA versus MSA

MSA has several advantages over PCA, some of which already became apparent. First, both methods can be used to study dimensionality but only MSA is able to test the psychometric properties of the scales that are found. Second, PCA will always result in components regardless of their theoretical usefulness. MSA on the other hand is based on assumptions about item-trait relations and can therefore support or reject a scale. Third, MSA is better able to handle categorical data because it avoids the difficulty factors of which PCA suffers (Wismeijer et al., 2008).

Dimensionality research is of great importance for the building of theory and it is therefore that this study will apply both PCA and MSA on the VRS. Using these two different methods will allow for definite conclusions about the VRS's dimensionality.

3. METHODS

3.1 Participants

The original sample of Tinlsley Bowman and York (1989) consisted of 252 participants: 84 male students, 134 female students and 34 students that did not indicate their gender. Ages ranged from 18 to 42 with a median of 19. 15 subjects were not deemed fit for analysis and were therefore deleted from the dataset which left a sample of 237. which was divided into two subsamples. The first half (118 subjects) was used for the exploratory part of the study. The second half of the sample (119 subjects) was used for confirmatory analysis on the scale(s) found using the Stoel procedure which is described below in paragraph 3.2.4. This method is called a 2-fold cross-validation (Stone, 1974). From here on out the first half of the sample will be referred to as training set and the second half of the sample will be called the test set.

3.2 Analytic strategy

3.2.1 PCA. To test whether the VRS contains a single factor or the three separate factors proposed by Barret and Tinsley (1977a), SPSS 18.0 was used to conduct a PCA. The Kaiser criterion and the scree test were used to determine the number of components. The single-factor solution was examined by looking at the corrected item-total correlation (Gerbing & Anderson, 1988). The three-factor solution was examined by means of direct oblimin rotation. The reasons for using an oblique rotation are explained in paragraph 2.2.2. Both the pattern matrix, which displays the correlations that can be compared to the Beta values of a multiple regression, and the structure matrix which show the correlations between the variable and the factor were examined and compared. When using orthogonal rotation the pattern matrix and the structure matrix are the same. However, when using an oblique rotation method the matrices differ. The pattern matrix shows the relationship between the variable and the factor with the influence of other variables partialled out. The structure matrix also shows the relationships between factors. Thus, when interpreting the data, both matrices are relevant. Interpreting these matrices can be tedious work, especially when there are numerous items. Because of this, Bredford's (Bredford, 1997) criteria were used: (a) correlations should be higher than .30 and (b) major loadings should be at least .20 higher than any cross-loading. In this manner, items with either low primary loadings or a variance spread over two or more factors will be excluded.

3.2.2 Scale Construction, Investigating Dimensionality. Classical clustering methods use a top-down procedure for creating scales. Deleting items from a scale until satisfaction with α is at its highest. In contrast MSA uses coefficient H_i which summarizes the strength between an item and the scale measuring the latent trait. The higher the H_i value, the better the item discriminates between high and low latent trait scores. A minimum value of H_j is usually set at .3 so that only items with minimum discriminative power are selected (Schuur, 2003). Coefficient H indicates the strength of the relationship between the total score and the latent trait. H uses information from H_i values and generally has the same minimum of .3 (Emons, Sijtsma, & Pedersen, 2010). Higher values of H indicate a more precise person ordering on the latent trait based on the total score. Generally three rules of thumb are used to indicate the strength of a scale. $.3 \leq H < .4$ indicates a weak scale. $.4 \leq H < .5$ indicates a medium scale. If $H \geq .5$ it means a strong scale (Wismijer et al., 2008).

3.2.3 Automated Item Selection Procedure. For selecting clusters of items, this study used a form of automated item selection procedure (AISP; Sijtsma & Molenaar, 1995). This procedure works in an exploratory way, finding clusters that each measure a different construct. This procedure is included in the software package MSP5 for Windows (also called MSP; Mokken scale analysis for polytomous items) and is executed as follows. In the first step the researcher appoints a value to *c.* This value articulates the minimum level coefficient H should have in each cluster. In MSP the c has a default value of .3. This can be increased for stricter selection or reduced for a less restrictive selection process. If the lowerbound is set to $c = 1$, the item selection procedure is looking for a perfect

Guttman scale (Hemker, Sijtsma, & Molenaar, 1995). The first cluster is formed by selecting the two items that share the highest significant *H* value. MSP then starts adding items on the basis of their H_i value in accordance to the items already in the scale. When the first cluster has formed and there are still items left unselected, the program starts to form another cluster in the same manner as the first. This process continues until there are no items left unselected or when the remaining items do not contribute to any of the clusters (Emmons, Sijtmsa, & Pedersen, 2010). Hemker, Sijtsma and Molenaar (1995) suggest that the researcher should run the AISP several consecutive times with increasing values of *c,* starting at 0.0 and ending at .55, with steps of .05. When used in this fashion, the procedure helps to spot dimensionality as follows. When the data is multidimensional, increasing *c* values will show three stages: (1) Almost all the items are all included in one scale; (2) two or more scales are formed; and (3) the scales from step two become smaller as strictness increases. Unidimensional data treated in the same manner will show the following three stages: (1) most or all of the items are in one scale; (2) a single smaller scale is found; and (3) one or a few small scales are found and items are being rejected (Hemker, Sijtsma & Molenaar, 1995). For multidimensional data the second stage should be regarded as final. Scales found in this stage should be treated as unidimensional scales. For the unidimensional data the results from the first stage should be considered as final (Hemker, Sijtsma, & Molenaar, 1995). Van Abswoude, Hemker, Vermunt and Van der Ark (2004) concluded that the AISP procedure included in MSP has an important drawback. They argue that using this procedure can obscure the true dimensionality of a scale. Through an example they explain that the first pair of items that is selected determines the next inclusions. That way, once an item is included into a cluster, it can no long be included in another cluster where it might have found a better fit. This problem partly arises because the AISP is solely data-driven and does not allow room for theoretical expertise. Sijtsma and Molenaar (2002) also cautioned researchers to think before letting the computer do all the work. In other words, theoretical consideration is important. Taking these cautions to heart, the next section will explain how this study circumvented the risks that come with purely data-driven analysis.

3.2.4 Creating scales with theory in mind. Imagine that the AISP was used to check the dimensionality of an item pool of 50. A single factor solution was produced which resulted in a scale of 35 items. This scale may in fact not be the best possible scale because it was entirely based on the first two-item cluster. The procedure does not investigate all the possible subsets of items as a potential best scale (Schuur, 2011). Imagine that the scale mentioned above is based on the items 6 and 9 because they shared the highest possible H value. Next, items 21, 24, 26 and 29 are added to the scale. Items 40 and 42 are not included in the final scale but closer examination shows that they have a theoretical similarity to items 6 and 9. Even when they are put together it results in a scale with satisfying psychometric properties. The only reason they were not viable for inclusion is because they did not combine well with items 21, 24, 26 and 29. This example shows that the purely data-driven AISP might not offer the best possible solution. Luckily MSP allows the user to specify a startset of items. This way scales can be formed around a theoretically solid base while retaining strong psychometric properties. This led Stoel (2012) to slightly adapt the AISP. Instead of letting the procedure run its purely data-driven course, Stoel suggested that the researcher should intervene. The clustering method disregards semantic similarities between items. The researcher with his theoretical prowess does not. Therefore it is important that the researcher imposes his expertise on the clustering procedure. After the first AISP is run with a lowerbound c value of .3 the researcher looks at the items with the highest H_i values. From these items, the researcher selects an itemset that is semantically cohesive and practically and or theoretically distinguishable. The AISP is run again but now with the selected items as a startset and the lowest H_i of this set as the lowerbound c. This restriction on the c value will prevent items that lower the homogeneity of the scale from being included. After the next run of the AISP the researcher again looks at the H_i values and checks if any other strong items fit in theoretically with the former starting set. If this is the case, these items will also be included in the startset and the AISP will be executed again. This continues until no more items are able to increase the homogeneity while adding semantic value. This procedure aims at finding a balance between the solely data-driven approaches and processes where the theory is (too) dominant. The current study followed the steps that were described above in order to create homogenous scales. Once a scale was found the items constituting this scale were removed from the item pool and the procedure was executed again. This was done until no more scales were found.

3.2.5 Confirmatory MSA. MSA can also be used in a confirmatory fashion. This study will utilize this in two ways. First, in order to test if the VRS functions under a one-factor model that assumes that the 40 items measure the same factor, all the 40 items were run through a test procedure using MSP 5.0 (Molenaar & Sijtsma, 2000). This test shows the item H values as well as the total H value. The default lower bound for c is 0.3 was adopted. Second, with the other half of the sample the scales that were found in the exploratory phase will be run through the test procedure in order to draw definite conclusions about their psychometric properties.

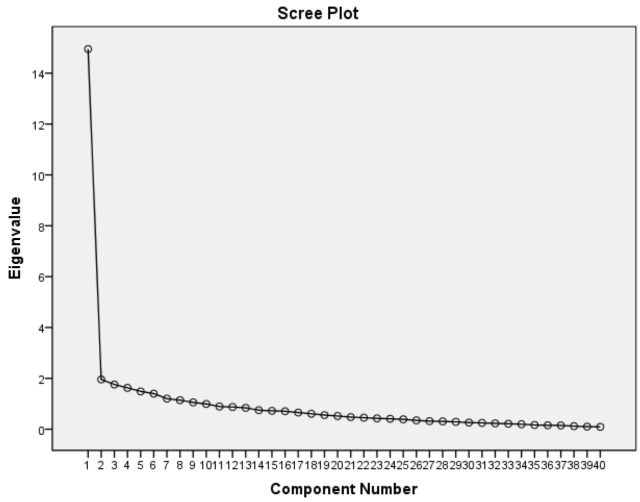
4. RESULTS

4.1 PCA on the VRS

PCA resulted in 9 eigenvalues greater than 1. The first component displayed an eigenvalue of 14.9 and explained 37.3% of the total variance. The following components had eigenvalues varying from the second component's 1.9 to the ninth component's 1.05. Total variance explained varied from 4.8% to 2.6%. The scree plot, which is depicted in figure 2, shows a sharp bend after the first factor, supporting a one factor solution. The item-total correlations are displayed in table 3. Only items 23, 33 and 1 show correlations lower than .30. Further analysis of these items showed that they were faulty rather than measuring a separate component. To test the null hypothesis, a three-factor solution was produced using oblimin rotation. Following the criteria of Bredford (1997), the structure matrix and pattern matrix showed similar results. The first component seems to measure some sort of knowledge-of-self concerning interests, capabilities and values in relation to career plans. The second component rather seems to measures a general self-knowledge. There was no reason to conclude that there was a third component. Table 3 further shows that there are a lot of items with major cross-loadings. Hence it is not surprising that the component correlation matrix showed that the first two factors correlated .54. Determining which items belong to which factor is better suited for MSA. Before turning to this specific calibration process, MSA was used to confirm if the VRS could be interpreted as a scale consisting of a single factor.

FIGURE 2

Scree plot of the VRS



4.2 Confirmatory MSA on the VRS

A confirmatory MSA showed that the *H* value for the final scale was .35, indicating a weak scale. Table 3 shows that items 1, 3, 12, 22, 23, 30, 31, 33, 36, 37, and 38 had unsatisfying *H* values. This indicates that these items do not discriminate appropriately on the underlying latent trait. Item 1 and 33 both had a negative *H* with one of the items in the scale. Most of the other items that did not satisfy the lower bound c also showed high cross-loadings in the PCA and did not seem to belong exclusively to any of the two components. Items 12, 37 and 38 did not satisfy the lower bound *c* but exclusively load on the second factor (according to the PCA). Note that the highest *H* values also showed the highest item-total correlations.

4.3 Exploratory AISP

The normal search procedure of MSP, executed with a lowebound *c* value of .3, resulted in a scale of 33 items with a total *H* value of .45 indicating a medium scale. The reliability coefficient Rho equaled .96. Items 1 and 33 were excluded due to negative *H* values with one of the items in the scale. Items 12, 23, 31, 36 and 38 were excluded because they did not meet the lower bound *c* value of .3. This was also the case in the confirmatory MSA. The search procedure did not yield any usable secondary scales.

4.4 Stoel Procedure

Running the AISP at .30 resulted in a scale of 33 items with a total H value of .45 and a RHO that equaled .96. The following items displayed the highest significantly positive H_i values and were considered semantically cohesive:

	Item	H_i
4	I'm really not sure of my occupational interests.	.52
9	I know my own values well enough to make a career decision right now.	.55
18	I feel confident that my career plans match my personality, interests (…).	.55

These items were taken as a startset for the subsequent AISP. The lowerbound value for the first run was determined by the lowest H_i of the selected items: .52. The results from this second EMSA let items 13, 11 and 27 to be included in the next startset. Items 13 and 27 shared the lowest H_i value of .62. Consequently the next AISP was run with a lowerbound c value of .62. The only items found in the scale that were not part of the startset were items 2 and 40:

	Item	H_i
2	I just can't make up my mind what type of work I'm cut out for.	.63
40	When it comes to choosing a college major or an eventual career, I'm really up in the air.	.66

TABLE 3
Corrected Item-Total correlations, Pattern and Structure coefficients after Oblimin Rotation of three components , and Item H values

	Item	Corrected Item-Total Correlation	Pattern Coefficients			Structure Coefficients			Item *H*
			Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	
4	I'm really not sure of my occupational interests.	.72	.88*	-.05	-1.2	.83*	.40	-.03	.45
27	I have only a foggy idea of what I'm interested in.	.72	.87*	-.08	-.03	.83*	.39	.05	.46
11	I know pretty much what I'm looking for in a college major and a career.	.74	.87*	-.04	-.02	.84*	.42	.07	.46
13	I haven't got the faintest idea what type of job (or college major) I'm looking for.	.64	.84*	-.15	.72	.77*	.32	.15	.44
2	I just can't make up my mind what type of work I'm cut out for.	.70	.76*	.06	-.21	.78*	.45	-.11	.44
18	I feel confident that my career plans match my personality, interests, etc.	.76	.72*	.13	.04	.79*	.53	.14	.48
40	I When it comes to choosing a college major or an eventual career, I'm really up in the air.	.74	.71*	.11	.01	.77*	.50	.10	.46
9	I know my own values well enough to make a career decision right now.	.76	.67*	.18	.12	.78*	.56	.22	.48
30	I can easily name three types of occupations in which I would feel satisfied.	.45	.63*	-.18	.25	.56*	.19	.30	.29
8	I'm not certain about what kind of job environment I'd be really happy in.	.60	.62*	.10	-.17	.66*	.42	-.09	.37
16	I'm very aware of my own values and how they will influence my choice of a career.	.52	.51*	-.03	.30	.57*	.30	.36	.34
7	I have a clear idea of my own needs and desires with respect to a career.	.64	.53*	.19	.13	.65	.50	.22	.41
39	I know enough about my interests and abilities (…) to predict what career I will be in five years from now.	.64	.49	.30	-.05	.65	.56	.03	.41
6	I'm not sure of what abilities I have that I can build a career around.	.66	.46	.39	-.27	.64	.61	-.17	.42
20	I just don't know if I have the traits that some lines of work require.	.73	.45	.44	-.14	.68	.67	-.03	.47
21	I have a real clear picture of my work-related attributes and characteristics.	.69	.45	.33	.11	.65	.59	.21	.45
15	With respect to the (…) things which would be important for a career, I don't know where my abilities lie.	.69	.43	.40	-.09	.64	.62	.01	.44
24	I've had a lot of different work experiences and I've learned what I need and want in a career.	.64	.41	.27	.32	.60	.54	.41	.41
19	If I had a clearer idea of what I'm like (…), I'd be able to make a decision about a major or a career.	.59	.39	.37	-.28	.56	.54	.19	.37
29	I don't know my values with respect to careers as well as I would like to.	.60	.39	.35	-.08	.57	.56	.01	.39

TABLE 3
(Continued)

	Item	Corrected Item-Total Correlation	Pattern Coefficients			Structure Coefficients			Item *H*
			Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	
28	I can easily think of 5 adjectives which I am certain would describe my (…) characteristics.	.58	.36	.23	.33	.52	.47	.40	.37
31	I have just a hazy notion of what "payoffs" or rewards I'm looking for in a career.	.37	.25	.14	.17	.35	.30	.22	.23
25	I really don't know myself as well as most other people my age.	.58	-.04	.76*	.18	.39	.76*	.27	.38
35	I would have some problems answering the question, "What sort of person are you?"	.63	-.01	.75*	.16	.43	.77*	.25	.35
12	If somebody was to describe me in terms of (…) I'd have trouble deciding if the description was accurate.	.40	-.03	.62*	-.18	.27	.57*	-.11	.26
26	I am certain that my knowledge of my own interests, abilities, etc., is accurate	.63	.15	.61*	.06	.50	.70*	.16	.41
32	I am as certain of what I'm like and what I have to offer to the world of work as anybody else my age.	.58	.05	.60*	.29	.42	.67*	.37	.38
37	I honestly feel that I don't need any counseling in making my future career plans.	.45	.13	.49*	-.22	.38	.53	-.14	.03
17	My weak (…) is that I just don't know myself well enough yet to be able to make a good career decision.	.72	.41	.48	-.14	.66	.69	-.03	.46
5	I know myself well enough to know what kind of job fits me.	.66	.30	.48	-.21	.57	.65	.11	.42
34	I don't know if I have the right personality for the kind of work I'm considering.	.55	.26	.48*	-.29	.48	.58	-.20	.35
38	I can't make up my mind whether I have the "drive" necessary to go to graduate or professional school.	.32	-.06	.47*	.05	.19	.44	.11	.20
10	I just can't put my finger on the best way to describe myself.	.52	.23	.41	-.05	.45	.53	.02	.33
36	If I had to choose between a business job and a people-helping job I would be able to decide very easily.	.32	.10	.25	.21	.27	.34	.26	.21
33	There are several occupations in which I know I would not fit.	-.09	.07	-.17	-.16	-.04	-.15	-.17	-.07
3	My past work experiences have taught me a lot about myself.	.45	.33	.10	.43	.44	.35	.48	.29
23	If I was told that my main requirement in a job was security, I wouldn't know for sure how accurate that was.	.24	.11	.08	.41*	.21	.20	.44*	.16
22	If someone asked me to describe my vocational strengths, I wouldn't know where to start.	.44	.28	.15	.33	.41	.35	.38	.28
4	On the basis of my past experience, I have a real clear picture of what kind of person I am.	.53	.28	.27	.32	.47	.47	.39	.35
1	I think I'm at the same point as most other people my age in terms of planning a future career.	-.04	-.03	-.05	.28*	-.03	-.04	.27*	-.02

Due to the H_i values and the mismatching content of item 2, only item 40 was admitted to the final scale which is displayed below in table 4. The items in the scale all pertain to the clarity of one's occupational and or educational values and interests.

TABLE 4
Final scale First run Stoel Procedure

	Item	Mean	H_i
9	I know my own values well enough to make a career decision right now.	3.26	.68
11	I know pretty much what I'm looking for in a college major and a career.	3.61	.73
40	When it comes to choosing a college major or an eventual career, I'm really up in the air.	3.64	.67
4	I'm really not sure of my occupational interests.	3.67	.68
18	I feel confident that my career plans match my personality, interests, etc.	3.78	.67
27	I have only a foggy idea of what I'm interested in.	3.94	.71
13	I haven't got the faintest idea what type of job (or college major) I'm looking for.	4.27	.70
Scalability coefficient H			.69
Reliability coefficient ρ			.93

When the decision was made to label this as the final scale, the items constituting this scale were removed from the pool and the procedure was repeated with the remaining items in search of a second scale. The first AISP with a lowerbound c value of .3 resulted in a scale of 25 items with an H value of .42 and RHO equaling .94. The same protocol applied when constructing the first scale was followed to form the second scale which is displayed below. The startset was formed by items 20 and 21. Rather than measuring the clarity and certainty of vocational or educational interest, these items seemed to measure clarity and certainty about one's abilities needed in a future career.

	Item	H_i
20	I just don't know if I have the traits that some lines of work require.	.53
21	I have a real clear picture of my work-related attributes and characteristics.	.50

With these items as a startset and a lowerbound c value of .50, the next AISP was run. Items 39, 15 and 2 showed the highest H_i values as well as semantic relevance. Item 17 showed a high H_i value as well but did not seem to match the content of the other items and was therefore not included in the next startset.

	Item	H_i
2	I just can't make up my mind what type of work I'm cut out for.	.56
15	With respect to (...) what would be important for a career, I don't know where my abilities lie.	.56
39	I know enough about my interests (...) to be able to predict what career I will be in five years from now.	.55
17	My weak point (...) is that I just don't know myself well enough yet to (...) to make a good career decision.	.60

Note that item 2 was also a candidate for the first scale but looking at the content closely reveals that it fits better in this second scale. 'Being cut out for something' is interpreted as having the right skillset or abilities rather than having a matching interest or personality. The next AISP was run with a lowerbound c value of .55. Only item 6 was deemed relevant to the construct and was therefore included in the scale. Items 5 and 17 both had acceptable H_i values but were not included due to a mismatch in content.

	Item	H_i
6	I'm not sure of what abilities I have that I can build a career around.	.56
5	I know myself well enough to know what kind of job fits me.	.55

The final scale is shown in table 5.

TABLE 5
Final scale Stoel Procedure Second run

	Item	Mean	H_i
39	I know enough about my interests (...) to predict what career I will be in five years from now.	3.23	.53
20	I just don't know if I have the traits that some lines of work require.	3.25	.61
6	I'm not sure of what abilities I have that I can build a career around.	3.36	.58
15	With respect to (...) what would be important for a career, I don't know where my abilities lie.	3.49	.59
21	I have a real clear picture of my work-related attributes and characteristics.	3.53	.60
2	I just can't make up my mind what type of work I'm cut out for.	3.66	.58
Scalability coefficient H			.58
Reliability coefficient ρ			.87

The items constituting the second scale were removed and the Stoel procedure was executed again. The first AISP resulted in a scale of 19 items with a total H value of .39 and a RHO of .91. Items 25, 26 and 35 showed the highest H_i values and were semantically cohesive in that they all pertained to a general self-knowledge.

	Item	H_i
25	I really don't know myself as well as most other people my age.	.45
26	I am certain that my knowledge of my own interests, abilities, etc., is accurate	.45
35	I would have some problems answering the question, "What sort of person are you?"	.45

The next run of the AISP showed that only item 32 was a viable candidate to be included in this scale. The final scale is depicted below in table 6.

TABLE 6
Final scale Stoel Procedure Third run

	Item	Mean	H_i
35	I would have some problems answering the question, "What sort of person are you?"	3.47	.60
32	I am as certain of what I'm like and what I have to offer to the world of work as anybody else my age.	3.57	.56
26	I am certain that my knowledge of my own interests, abilities, etc., is accurate	3.80	.61
25	I really don't know myself as well as most other people my age.	3.92	.67
Scalability coefficient H			.61
Reliability coefficient ρ			.87

4.5 Confirmatory MSA on the new scales

In this section the scales that were found in the exploratory phase will be put through the test procedure provided by MSP 5.0 using the second half of the sample. The results are displayed in the tables below. The column under A shows the values found in the exploratory phase. The values displayed in the B column are the values found in the confirmatory analysis on the test set. The first scale measures clarity and certainty of interests and values related to career and education. An adequate label for this scale would be: *Crystallization of Vocational Interests and values*. The confirmatory analysis shows an H value of .62 and a reliability of .91, indicating an excellent scale. Results are shown in table 7.

TABLE 7
Confirmatory MSA on the Crystallization of Interests and Values Scale

	Item	Mean	H_{iA}	H_{iB}
9	I know my own values well enough to make a career decision right now.	3.20	.68	.61
11	I know pretty much what I'm looking for in a college major and a career.	3.52	.73	.64
40	When it comes to choosing a college major or an eventual career, I'm really up in the air.	3.72	.67	.58
4	I'm really not sure of my occupational interests.	3.64	.68	.61
18	I feel confident that my career plans match my personality, interests, etc.	3.72	.67	.56
27	I have only a foggy idea of what I'm interested in.	3.89	.71	.67
13	I haven't got the faintest idea what type of job (or college major) I'm looking for.	4.25	.70	.67
Scalability coefficient H			.69	.62
Reliability coefficient ρ			.93	.91

The second scale consists of items that pertain to crystallization of the knowledge of one's vocationally relevant abilities concerning a future career. *Crystallization of Vocational Abilities* would be an appropriate label for this scale. A total H value of .49 and a reliability of .84 indicate a strong scale. Results are shown in table 8.

TABLE 8
Confirmatory MSA on the Crystallization of Vocational Abilities Scale

	Item	Mean	H_{iA}	H_{iB}
39	I know enough about my interests (...) to be able to predict what career I will be in five years from now.	3.22	.53	.48
20	I just don't know if I have the traits that some lines of work require.	3.21	.61	.44
6	I'm not sure of what abilities I have that I can build a career around.	3.42	.58	.52
15	With respect to (...) which would be important for a career, I don't know where my abilities lie.	3.40	.59	.59
21	I have a real clear picture of my work-related attributes and characteristics	3.45	.60	.49
2	I just can't make up my mind what type of work I'm cut out for.	3.62	.58	.44
Scalability coefficient H			.58	.49
Reliability coefficient ρ			.87	.84

The third scale measures general self-knowledge. An appropriate label for this scale would be: *Crystallization of the Self-concept.*

TABLE 9
Confirmatory MSA on the Crystallization of the Self-concept Scale

	Item	Mean	H_{iA}	H_{iB}
35	I would have some problems answering the question, "What sort of person are you?"	3.51	.60	.49
32	I am as certain of what I'm like and what I have to offer to the world of work as anybody else my age.	3.66	.56	.46
26	I am certain that my knowledge of my own interests, abilities, etc., is accurate	3.69	.61	.47
25	I really don't know myself as well as most other people my age.	3.82	.67	.56
Scalability coefficient H			.61	.50
Reliability coefficient ρ			.87	.79

5. DISCUSSION

The main purpose of this study was to examine the dimensionality of the VRS. In doing so, this study compared the results from PCA to the results from a modern technique from IRT, namely, MSA. An addition was made to the standard procedure of carrying out a dimensionality check using the AISP. Rather than letting only the data speak, the procedure was guided by the researcher in order to find a balance between strong psychometric properties and theoretical utility.

First a PCA was carried out on the first half of the sample. Results regarding dimensionality were ambiguous at first glance. The corrected item-total correlation supported a single-factor solution. To test the three-factor solution a PCA was executed using direct oblimin rotation. This resulted in two interpretable components which by itself suggest multidimensionality. However, the high correlation between these factors (.54) and the high corrected item-total correlations strongly suggest the VRS to be unidimensional. Next the VRS was run through a test procedure provided by MSP to conclude if the VRS was able to function as a unidimensional scale. The total H value of the scale was .35 which indicates a weak scale. The results also showed that 11 items were not scalable and therefore these items were removed from the scale. An exploratory search procedure with a lowerbound c value of .3 indicated a medium scale of 33 items with an H of .45 and a reliability of RHO equaling .96. 7 items were deemed unscalable. Thus the confirmatory and exploratory MSA also suggested a single-factor solution. A few things should be noted however. The PCA did produce two separate and interpretable components. Both confirmatory and exploratory MSA showed that some items were not scalable. Also the PCA showed major cross-loadings on almost half the items. Whether the VRS is or is not unidimensional, it is flawed to some extent. The above suggest there are redundant as well as defective items. The methods mentioned were not able to draw a singular, definite conclusion about the dimensionality of the VRS.

The adapted version of the AISP, the Stoel procedure, was able to make distinctive scales with solid psychometric properties and in doing so, getting rid of redundant and flawed items. The first scale that was produced resembled the first component found in the PCA. Although, instead of the 12 items that comprised the first component, this scale consisted of 7 items. The items with the highest factor loadings were 4, 27, 11, 2, 13, 9, 18 and 40. With the exception of item 2 ('I just don't know what type of work I'm cut out for), these items formed the first scale in the Stoel procedure. During the first run of the Stoel procedure, the H_i of item 2 did satisfy the need for homogeneity but looking closely at the content it became less viable for inclusion. 'Being cut out for something' seems to tap into the construct of ability rather than interests and values. When you are not able to do something, you are not cut out for it. Admittedly this is up for debate. Item 2 however found its place in the second scale, measuring crystallization of knowledge about one's vocational abilities. Like the first scale, this group of items also showed a strong measure of homogeneity and semantic cohesion. This scale did not show any resemblance with the second component found in the PCA. However, the third scale consisted of items that also could be found in the second component produced by the PCA. The scale was formed by items 25, 26, 32 and 35, measuring general knowledge about the self. This is why the scale was labeled as *crystallization of the self-concept.* The second component also included item 12, 34 37 and 38. When looking at it from a purely theoretical viewpoint, item 12 ('If somebody was to describe me in terms of my personality, interests, etc., I'd have trouble deciding if the description was accurate.') was a good candidate for inclusion. It did however not meet the restrictions put up by the lowerbound *c.*

When we move beyond the comparison of MSA and PCA, what implications do the results of the former have on the theory of vocational self-concept crystallization? Indeed, three scales were found but do they reflect clarity, certainty and structure? The simple answer is no. These results offer no reason to believe that there can be made a distinction between a factor measuring clarity and a factor tapping into the construct of certainty, let alone structure. The factor analysis of Tinsley, Bowman and York (1989) neither found an indication that structure might form a separate factor. They did conclude however that the VRS was loading on the first two factors they found: clarity and certainty. Like mentioned before, this study included four questionnaires thus making it hard to draw definite conclusions about the dimensionality of the VRS. Do the results of this study suggest that the research on the concept of vocational self-concept crystallization should be abandoned? On the contrary, when looking at the scales that were found from a language perspective, all the scales seem to contain items that ask for clarity as well as certainty. This suggests that all scales measure crystallization to some extent. These results do not reject the concept of vocational self-concept crystallization, they only sharpen it. The concept of vocational self-concept crystallization revolves around the knowledge of self in regard to vocational relevant values, interests, attitudes, needs and abilities (Barret & Tinsley, 1977a). Although Barret and Tinsley (1977a) put clarity certainty and structure under the singular term vocational self-concept crystallization, it is unclear if they believed these metadimensions to be psychometrically separable. It turns out they are not. They did however overlook the usefulness of dividing crystallization in terms of abilities, interest, values and needs. A person can have crystallized knowledge about his own interests and values but this crystallization might be lacking when it comes to abilities or needs. This study produced one scale that measures the former and a second scale that measures the latter. It can be useful to separate between persons on these two constructs. It can be highly interesting to study which form of crystallization is more important when choosing a career. Also a third scale was found which seems to measure crystallization of general self-knowledge. Future research should focus on the causal links between these three concepts and how each them relates to career development.

By using MSA it was possible to create three scales with strong psychometric properties. These properties were tested by doing a confirmatory MSA on the second half of the sample and it was concluded that the strong psychometric backbone was consistent. This research provides three strong scales that are able to measure vocational self-concept crystallization on three separate levels of interest: interests and values, abilities and general self-knowledge. Note however that the definition of Barret and Tinsley (1977a) also mentions *needs*. Although there were some items in the VRS that asked participants about their needs within a career (e.g. item 23 and 31), these items were not able to form a solid scale. To get a complete and clear picture about a person's vocational self-concept crystallization it is important that also the crystallization of one's needs can be measured. If this is also done by a form of IRT, preferably MSA, it becomes possible for vocational counselors and assessment centers to order persons on each of the scales to see where there is work to be done.

Another suggestion for future research comes from a methodological issue. Although this study compared the results of PCA to the results of the Stoel procedure, the latter was not compared to the traditional dimensionality check provided by Hemker, Sijtsma and Molenaar (1995) in which the lowerbound c is increased from .0 to .55 with steps of .05. For future research, it is interesting to examine how the results from a purely data-driven procedure differ from the results of a procedure where there is an intervention of the researcher. Wismeijer et al., 2008 concluded that MSA is better suited for checking dimensionality and finding scales than PCA. They also concluded that the two methods are complementary. When examining the dimensionality of the Self-Concealment Scale (Larson & Chastain's 1990) and thereby comparing MSA and PCA, they found that the dimensionality check produced similar results. They also concluded that the final scale produced by MSA consisted of fewer items than the component suggested by PCA. The current study also showed that final scales resulting from MSA were comprised by a fewer number of items than comparable components from PCA. Where Wismeijer et al. (2008) found similar results of the dimensionality check, this study showed that MSA, used in a expertise-driven fashion, resulted in three factors while PCA only suggested two.

The current study has provided a boost for research on a theoretical level as well as an impulse for research on methodological issues. The concept of vocational self-concept crystallization was sharpened and a new way of executing the standard AISP was presented which may spark further research into the perfecting of MSA.

REFERENCES

Barret, T. C. , & Tinsley, H. E. A. (1977). Measuring vocational self-concept crystallization. *Journal of Vocational Behavior, 11*, 305-313.

Barret, T. C. , & Tinsley, H. E. A. (1977). Vocational self-concept crystallization and vocational indecision. *Journal of Counseling Psychology, 24(4),* 301-307.

Baruch, Y. (2004) *Managing careers: Theory and practice.* Essex : Pearson Education Limited.

Bedord, A. (1997) On Clark-Watons's tripartite model of anxiety and depression. *Psychological Reports, 80,* 125-126.

Bernstein, I. H. , & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105(3),* 467-477.

Chang, C., & Reeve, B. B. (2005). Item response theory and its applications to patient-reported outcomes measurement. *Evaluation & Health Professions, 28(3),* 264-284.

Conway, J. M., & Huffcut, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6(2),* 147-168.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: Guilford Press.

Defillippi, R. J., & Arthur, M.B. (1994). The boundaryless career: A competency based perspective. *Journal of Organizational Behavior, 15*, 307-324.

Domingos, P. (1999). The role of Occam's Razor in knowledge discovery. *Data Mining and Knowledge Discovery, 3*, 409-425.

Eby, L. T., Butts, M., & Lockwoord, A. (2003). Predictors of success in the era of the boundaryless career. *Journal of Organizational Behavior,* 24, 689-708.

Embertson, S. E., & Reise, S.P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Emmons, W. H. M., Sijtsma, K., & Pederson, S. S. (2010). Dimensionality of the hospital anxiety and depression Scale (HADS) in cardiac patients: Comparison of mokken scale analysis and factor analysis. *Assessment,*, -.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4(3),* 272-299.

Gerbing, D.W., Anderson, J.C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25(2),* 186-192.

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous mokken IRT model. *Applied Psychological Measurement, 19(4),* 337-352.

Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement, 25(3),* 211-220.

Lunnenborg, P. W. (1978). Sex and career decision-making styles. *Journal of Counseling Psychology*, *25(4)*, 299-305.

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *International Journal of Educational and Psychological Assessment, 1,* 1-11.

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response modeling. *Psychological Methods, 9(3),* 354-268.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* New York/Berlin: De Gruyter.

Noe, R. A., Noa, A. N., & Bachhuber, J. A. (1990). An investigation of the correlates of career motivation. *Journal of Vocational Behavior,* 37, 340-356.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment test.* Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Rouse, S. V., Finger, M. S., & Butcher, J. N. (1999). Advances in clinical personality measurement: An item response theory analysis of the MMPI-2 PSY-6 scales. *Journal of Personality Assessment, 72(2),* 282-307.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8(4),* 350-353.

Schuur, W. (2003). Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory. *Political Analysis, 11*, 139-163.

Schuur, W. (2011). *Ordinal item response theory.* Los Angeles: SAGE publications, Inc.

Sijtsma, K., Emons, W. H. M., Bouwmeester, S., Nyklíček, I., & Roorda, L. D. (2008). Nonparametric IRT analysis of quality-of-life scales and its application to the World Health Organization quality-of-life-scale (WHOQOL-brief). *Quality of Life Research, 17,* 275-290.

Sijtsma, K., & Molenaar, I. W. (2002). *An introduction to nonparametric item response theory.* Thousand Oaks: Sage Publications.

Stone, M. (1974). Cross-validatory choise and assessment of statistical predicitons. *Journal of the Royal Statistical Society, 36(2), 111-147.*

Super, D. E., Starishevsky, R., Martlin, N., & Jordaan, J. P. (1963). *Career development: Self-concept theory* (CEEB Research Monograph No. 4). New York: College Entrance Examination Board.

Taylor, M. S. (1985). The roles of occupational knowledge and vocational self-concept crystallization in students' school-to-work transition. *Journal of Counseling Psychology, 32(4)*, 539-550.

Tinsley, H. E. A., Bowman, S. L., & York, D. C. (1989). Career Decision Scale, My Vocational Situation, Vocational Rating Scale, and Decisional Rating Scale: Do they measure the same constructs? *Journal of Counseling Psychology, 36*, 115-120.

Tokar, D. M., Withrow, J. R., Hall, R. J., & Moradi, B. (2003). Psychological separation, attachment security, vocational self-concept crystallization, and career indecision: A structural equation analysis. *Journal of Counseling Psychology, 50*, 3-19.

van Abswoude, Van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement, 28(1),* 3-24.

Van Abswoude, A. A. H., Vermunt, J. K., Hemker, B. T. H. , Van der Ark, L. A. (2004). Mokken scale analysis using hierarchical clustering procedures. *Applied Psychological Measurement, 28(5),* 332-354.

Weng, Q., & McElroy, J. C. (2010). Vocational self-concept crystallization as a mediator of the relationship between career self-management and job decision effectiveness. *Journal of Vocational Behavior, 76*, 234-243.

Wismeijer, A. A. J. , Sijtsma, K. , Van Assen, M. A. L. M. , & Vingerhoets, A. J. J. M. (2008). A comparative study of the dimensionality of the self-concealment scale using principal components analysis and Mokken scale analysis. *Journal of Personality Assessment, 90(4),* 323-334.