

Masculinities Representations Inventory (MRI, English Version): A measure of gender
(re)presentation
Russell Luyt
University of Greenwich

Author Note

Russell Luyt; Department of Psychology, Social Work and Counselling; University of Greenwich.

This research was supported in part by grants from the Commonwealth Commission and Universities UK. Thanks go to Sarah Kempster, Rosemary Lobban, Viyada Phouang and Taryn van Niekerk for their research assistance; Daragh McDermott for his helpful advice; Ronald. F. Levant and James M. O'Neil for making their work accessible; and the University of Cape Town for providing access to research participants.

Correspondence concerning this article should be addressed to Russell Luyt, Department of Psychology, Social Work and Counselling; University of Greenwich, London, SE9 2UG, United Kingdom. Email: r.luyt@gre.ac.uk

Abstract

This paper introduces the Masculinities Representations Inventory (MRI), English version, as a multidimensional measure of gender (re)presentation. It provides structural, convergent and divergent validity, as well as reliability evidence, in support of its use among English speakers in South Africa. Principal components analysis, through means of a male student sample ($n = 319$), confirms the measure's construct multidimensionality. Three factors inform a 29 item total- and subscale measures including dominant Representations of Othering (Anti-Effeminacy and Homo-Negativity), Responsibility (Dependability and Success), and Control (Dominance and Toughness). Evidence of convergent validity is seen in predicted patterns of correlation between MRI scale scores and those of the Male Role Norms Inventory as well as Gender Role Conflict Scales. Evidence of divergent validity is apparent in nonsignificant correlations, in all but one case (Masculinity), with the Personal Attributes Questionnaire scale scores.

Keywords: Masculinities Representations Inventory (MRI), masculinity ideology, masculinity (re)presentation, measurement

Masculinities Representations Inventory (MRI): A measure of gender (re)presentation

This paper introduces the Masculinities Representations Inventory (MRI) as a multidimensional measure of gender (re)presentation and seeks to provide measurement validity evidence in support of its use among English speakers in South Africa. The MRI is designed to index socio-cultural group endorsement of dominant gender representations. The Inventory is conceptually firmly grounded within a social constructionist understanding of gender and emerges through the revision of the Male Attitude Norms Inventory-II (MANI-II) (Luyt, 2005). The revision of the MANI-II and subsequent development of the MRI was considered necessary for a number of reasons. Firstly, whilst the MANI-II was explicitly developed as a measure of ‘masculinity ideology/ies’, recent theoretical debate suggests the worth of developing measures of ‘masculinity (re)presentation’ (Luyt, 2013/2015). Secondly, insufficient evidence exists in support of the MANI-II’s cross-cultural measurement validity in South Africa (Luyt, 2012). These reasons are discussed in turn below.

The Gender (Re)presentation Approach to Measurement

Luyt (2013/2015) argues that the gender (re)presentation approach to measurement offers an alternative to the ‘gender orientation’ and ‘gender ideology/ies’ approach. These latter two approaches, first described by Thompson, Pleck and Ferrera (1992), seek to measure gender as an individual phenomenon; either as individual traits or individual endorsement and internalization of social norms (Thompson & Pleck, 1986). That is to say, measures of gender orientation seek to assess individual attributes, characteristics or traits. These are believed to reflect real differences between men and women “which are more or less rooted in anatomy, physiology” or “early experience” (Constantinople, 1973, p. 390). Well-known gender orientation measures include the Bem Sex-Role Inventory (BSRI; Bem, 1974) and the Personal Attributes Questionnaire (PAQ; Spence, Helmreich & Stapp, 1974). Measures of gender ideology/ies, alternatively, seek to assess “an individual’s internalization

of cultural beliefs regarding gender roles” (Levant, Richmond, Cook, House & Aupont, 2007, p. 373). The Brannon Masculinity Scale (BMS; Brannon & Juni, 1984) represents an early gender example. The development of new [e.g., the Femininity Ideology Scale (FIS; Levant, Richmond, et al., 2007)] as well as the revision of existing measures [e.g., the Male Role Norms Inventory – Revised (MRNI-R; Levant, Smalley, et al., 2007)] points, however, toward a still fertile research field. Thompson and Bennett’s (2010) recent review of evolving measures of masculinity ideology/ies captures this well.

Notwithstanding the wide array of gender-related constructs, their varied classification (e.g., Kroska, 2000; McCreary, Newcomb & Sadava, 1998; Smiler, 2004; Smiler & Epstein, 2010,) and criticisms leveled at the distinction between gender orientation and gender ideology/ies (Thompson & Pleck, 1995), this classification holds advantages. It informs the appropriate use of instruments and aids the development of theoretically sound measures (Thompson, et al., 1992).

Yet, as is noted above, both the gender orientation and gender ideology/ies approach seek to measure gender as an individual phenomenon. Luyt (2013/2015) describes how these measures operationalize the evolving assumptions of gender role theory. It is argued that social constructionist critique problematizes this conceptualization and suggests that gender is better understood and measured as a social phenomenon.

Social constructionism (Beall, 1993; Burr, 1998) and gender role theory are not, as Pleck (1995) asserts, compatible theoretically. They assume profoundly different meta-theoretical positions. Such assertions are nonetheless fairly commonplace. This is seen in claims, for example, that the concepts of the ‘traditional male role’ and that of ‘hegemonic masculinity’ are congruent (e.g., Levant & Richmond, 2007). Yet Connell’s (1995) early definition of hegemonic masculinity identifies this as a pattern of gender *practice* in which care is taken to differentiate the concept and associated theorizing from that of gender role

theory (Connell, 1987, 1990, 1993). Such assertions appear to emanate from narrow understanding concerning the core assumptions of social constructionism. Levant, Richmond, et al. (2003, p. 92) exemplify this in their description of this theoretical perspective, as it applies to masculinities; namely “that there is no invariant masculinity (in the case of men) but rather there are ‘masculinities’ that vary according to the social context”.

Social constructionists, including Psychologists (e.g., Gergen, 1985) among their ranks, assert far more than the contextual variability of gender. It is claimed that people do not ‘have’ gender, for example, in the form of individual traits or internalized social norms. Rather we ‘do’ gender (West & Zimmerman, 1987). It is a situated social practice informed through social representation concerning what it means to be a man, women, or a member of any other potential sex category. Social representation of gender occurs at a cultural, discursive or ideological level and serves to structure relations of power in society. It is (re)produced symbolically, for example through language, and offers shared ways of understanding. Gender practice is made possible through such representation whilst at the same time (re)producing it. It occurs at an individual, interpersonal and institutional level. The notion of identity may be explained as a subject position adopted by individuals, relative to social representation, and strategically claimed through *situated* social practice (Burr, 1998; Edley, 2001; Marecek, Crawford & Popp, 2004).

Gender is therefore neither essential nor enduring. People may hypothetically adopt any gender practice within the unfolding course of interaction, irrespective of their assigned sex category, so long as these social practices are made available and meaningful through social representation. In everyday life such agency is limited by social norms and institutions that govern ‘appropriate’ gender practice. Yet social constructionism provides a conceptual framework through which to explain occasions in which individual gender practices and

associated identities appear inconsistent or contradictory (Burr, 1998; Edley, 2001; Marecek, Crawford & Popp, 2004); not only across context but also within situated interaction.

Social constructionism informs the gender (re)presentation approach to measurement. This approach seeks “to index the extent to which groups endorse dominant gender representations, which serve to legitimate and (re)produce unequal gender relations, and in so doing make specific gender subject positions available” (Luyt, 2013/2015, p. 222). It, together with a critical review of gender measurement literature, suggests that:

...measures of gender (re)presentation should include five key characteristics: an emphasis on multiple masculinities, femininities and other gender practices (construct multidimensionality); a clear distinction between masculinity, femininity, and other gender-related concepts (construct independence); a focus on social or group level, as opposed to individual, phenomena; suitable evidence of measurement validity and theoretically appropriate content. (Luyt, 2013/2015, p. 220)

The importance of construct multidimensionality and independence is well-established (Constantinople, 1973; Morawski, 1987) and similarly emphasized in measures of gender ideology/ies (Thompson, et al., 1992; Thompson & Pleck, 1995). Multidimensional measurement facilitates nuanced understanding of gender attitudes by enabling us to explore variable endorsement of underlying dimensions. Moreover measuring gender constructs independently recognizes that although these constructs are often defined in opposition to one another (e.g., masculinity and femininity), this is not always the case. Some of their defining features may be shared.

Emphasis upon developing theoretically appropriate content is also common to measures of gender ideology/ies (Thompson, et al., 1992; Thompson & Pleck, 1995) and gender (re)presentation. The wording of items in both cases is similar and reflects some of their shared conceptual assumptions. Appropriate content includes the development of third

person statements based upon prescriptive norms. The use of plural (e.g., *masculinities* as opposed to masculinity) is commonplace as well. This underlines the assumption of construct multidimensionality, but also, in the case of gender (re)presentation, diversity in gender practice.

However, focus on gender as an exclusively social phenomenon is unique to the measurement of gender (re)presentation and thoroughly reflects social constructionist thinking. It is this that distinguishes the gender (re)presentation approach from the gender ideology/ies approach to measurement. Accounting for gender at an individual level is deemed incongruous when we consider it something people ‘do’ rather than ‘have’. People make sense of normative statements included in measures of gender (re)presentation in a contextually meaningful way. They are ‘doing gender’ when responding individually to such statements. These responses do not reflect some internal, pre-formed and stable mental state. Rather they point toward available systems of meaning, which are shared by other group members, and strategically deployed by them in order to make sense of their world and claim subject positions within the course of situated interaction. Analyzing data at the group or social level allows us to understand how normative or dominant gender representations are variably endorsed across and within socio-cultural groups, the potential functions such variations serve, and the subject positions that are made possible as a result (Luyt, 2013/2015).

Lastly, despite literature consistently placing importance on obtaining suitable measurement validity evidence in support of instrument use (Joint Committee on Standards for Educational Evaluation, 1999), such evidence is surprisingly lacking in published research (Hogan & Agnello, 2004). This is particularly the case in cross-cultural research. Gender measures that have been developed within a specific cultural context are often applied in contexts for which there is insufficient measurement validity evidence (Gibbons,

Hamby & Dennis, 1997; Smiler & Epstein, 2010). This problem is arguably exacerbated when authors do not clearly specify the cultures in, and for which, measures have been developed. This risks implying and therefore promoting an unsubstantiated universalism in gender attitudes, beliefs, experience, and ideas. Measures of gender (re)presentation reaffirm the importance of obtaining suitable evidence for the use of an instrument among a specified population where measurement equivalence is neither considered desirable nor necessarily possible at the outset.

The revision of the MANI-II and subsequent development of the MRI was considered necessary for two key reasons. The theoretical distinction between measures of gender ideology/ies as opposed to measures of gender representation suggests that critical revision of the MANI-II, which was explicitly developed as a measure of ‘masculinity ideology’, is beneficial. Moreover such revision is recommended because, as described below, insufficient evidence exists in support of the MANI-II’s cross-cultural measurement validity in South Africa.

Measurement Revision of the Male Attitude Norms Inventory-II (MANI-II)

The MRI emerges through the successive development and revision of the MANI (Luyt & Foster, 2001) and MANI-II (Luyt, 2005). The MANI was originally developed in order to explore men’s endorsement of traditional masculinity in South African gang culture. Although not developed specifically as such, it was informed by two existing measures of masculinity ideology/ies – the Male Role Norms Inventory (MRNI; Levant et al. 1992) and the MRNS (Thompson & Pleck, 1986). This was subsequently revised in the form of the MANI-II (Luyt, 2005) in order to offer a valid multidimensional measure of masculinity ideology/ies in South Africa. This seeks to assess the degree to which individuals support notions of masculinity that justify prevailing gender scripts and associated power

relationships. The MANI-II is therefore congruent with other available gender ideology/ies measures [e.g., the Male Role Norms Inventory – Revised (MRNI-R) and the MRNS]. It offers the only published and validated means of assessing masculinity ideology/ies in South Africa. In this respect Hearn (2010) notes that the MANI-II signifies a growing trend toward greater cultural sensitivity in gender measurement.

Evidence exists in support of the MANI-II's (English Version) measurement – content and construct-related – validity. Yet it has been suggested that future research should include not only a more representative sample but also enhance construct validity evidence through an assessment of divergent validity (Luyt, 2005). In addition, whilst Afrikaans and Xhosa versions of the Inventory were produced through back-translation (Brislin, 2000) of the original English version, they were merely assumed to be equivalent. Insufficient cross-cultural validity evidence exists in support of the MANI-II's use among Afrikaans and Xhosa speakers.

Luyt (2012) developed a framework for mixing methods in quantitative measurement development, validation and revision. The author considers how this framework may be applied in the cross-cultural validation and subsequent suggested revision of the MANI-II in the form of a case study. Qualitative and quantitative analyses were conducted in order to ascertain whether cross-cultural content and construct validity evidence supports the use of different language versions. In motivating the current paper, both qualitative and quantitative findings indicate that Afrikaans, English and Xhosa versions do not reveal satisfactory evidence in support of their cross-cultural measurement validity. They suggest the benefit of specific revisions to the MANI-II in order to enhance evidence in support of this. Currently, as noted above, evidence only exists in support of the MANI-II's (English Version) content and construct-related measurement validity.

In sum, insufficient evidence exists in support of the MANI-II's cross-cultural measurement validity in South Africa. This, together with recent theoretical debate concerning 'masculinity (re)presentation', underpins the MRI's development. This paper presents the ensuing revisions and seeks to provide measurement validity evidence in support of its use among English speakers in South Africa. Four hypotheses were formulated in order to examine this.

Hypotheses

Hypothesis 1. EFA will confirm construct multidimensionality. Resultant subscale scores will have significantly higher positive correlations with the total scale score than with one another.

Construct multidimensionality is, as has been discussed, a well-established characteristic of gender measures. It features as an important assumption in the measurement of gender (re)presentation and facilitates a multifaceted understanding of gender attitudes by enabling us to explore variable endorsement of underlying dimensions.

Hypothesis 2. MRI scale scores will demonstrate meaningful significant positive and negative correlation, indicating moderate to strong construct convergence, with those of the MRNI. The MRNI is a measure of gender ideology/ies. It is most closely theoretically aligned to the MRI as a measure of gender (re)presentation. That is to say, these measures share theoretical emphasis in assessing the endorsement of social norms, despite differing in examining these at an individual versus a social level. Secondly, they apply a number of common measurement assumptions, such as construct multidimensionality and independence. And thirdly, the development of the MRI took place through the revision of the MANI-II, which has demonstrated a high positive correlation (South Africa, $n = 339$; $r = .84$, $p < .001$) with the MRNI (Luyt, 2005). These measures will therefore demonstrate the strongest

convergence. Whilst use of the MRNI-R (Levant, Smalley, et al., 2007) may also have been appropriate, the measure was not available at the time of data collection.

Hypothesis 3. MRI scale scores will demonstrate meaningful significant positive correlation, indicating low to moderate construct convergence, with those of the GRCS. The GRCS is a measure of gender role conflict. It holds some conceptual assumptions in common with the MRI, but is not as closely theoretically related as compared to measures of gender ideology/ies, and will therefore not demonstrate as strong convergence.

The GRCS may be classified as a measure of gender role strain (Thompson & Pleck, 1995), gender role stress (McCreary, et al., 1998), or strain (Smiler, 2004; Smiler & Epstein, 2010). The Masculine Gender Role Stress Scale (MGRSS) (Eisler & Skidmore, 1987) exists as another well-known example of this class of measure. Although they are similar to measures of gender ideology/ies in their assumption that gender is determined through social norms but ultimately internalized by individuals, they differ in their focus on how men experience their gender in terms of the degree of individual conflict or stress produced by norms of masculinity (Thompson & Pleck, 1995). Low to moderate, but positive correlations between GRCS scores and those of masculinity ideology/ies measures, support the view that they assess related but somewhat different constructs (O'Neil, 2008). It is similarly hypothesized (i.e., Hypothesis 3) that GRCS scores will exhibit low to moderate positive correlations with those of the MRI. These measures share an emphasis on social norms. Yet they differ in two key ways. The GRCS seeks to measure men's experience of their gender whilst the MRI seeks to assess the endorsement of dominant representations of masculinity. These constructs also differ in that they correspondingly consider gender at an individual versus a social level.

Hypothesis 4. It is hypothesized (i.e., Hypothesis 4) that PAQ-M scores will demonstrate nonsignificant correlations with those of the MRI as these assess the distinct

constructs of gender orientation versus gender (re)presentation in turn. As discussed, these measures differ in assessing individual traits as opposed to the endorsement of social norms, as well as examining these at an individual versus a social level. They also apply a number of disparate measurement assumptions concerning construct dimensionality and independence. The PAQ-M and the MRI will demonstrate the weakest relationship as they assess the most theoretically distinct constructs.

Method

Sample

Three hundred and twenty four male students from a large university in Cape Town, South Africa, agreed to complete the validation questionnaire. The majority of these were undergraduates (93.5%; $n = 303$) as compared to postgraduates (5.9%; $n = 19$). The mean age therefore remained low (20.6 years old; $n = 324$) despite a wide age range of between 18 and 38 years old. Similarly most students were single (84.6%, $n = 274$), a small proportion indicated having a partner (15.1%, $n = 49$), whilst none were either married or divorced. The distribution of participants across population group also appeared skewed, although arguably not markedly so, given that the questionnaire was completed in English. The majority of respondents classified themselves, in accordance with standard population group descriptors as applied by the national statistical service of South Africa, as White (56.2%; $n = 182$) followed by Black (17.9%, $n = 58$); Colored (14.2%, $n = 46$); Indian/Asian (8%, $n = 26$); and other (3.1%, $n = 10$). This sample should not be seen as representative of the male South African population (Statistics South Africa, 2013). Participants were recruited from a privileged educational environment and disproportionately represented the views of young adults. Results should be understood from within these constraints. Following listwise deletion of missing data, 319 cases were retained for primary analysis.

Measures

Validation questionnaires were developed. Each included a demographics page and two instruments. The demographics page required participants to indicate their age, gender, home language, marital status, population group, and whether they were either undergraduates or postgraduates. Questionnaires consisted of the MRI and one of three other measures including the GRCS, MRNI or PAQ. Six different versions of the questionnaire were produced so as to minimize both fatigue effects, and to ensure that instruments were evenly counter-balanced, thereby reducing the possibility of order effects.

MRI. The MRI is a 48 item measure of masculinity (re)presentation. It belongs to the broader class of measures which assess gender (re)presentation. The measure was developed in order to operationalize the revised systematized concept, including six theoretical dimensions and their underlying concepts. Eight items were developed in order to operationalize each of the dimensions as informed by a mixed method framework in quantitative measurement development, validation and revision (Luyt, 2012). This included both qualitative analysis of six focus group discussions and quantitative analysis of survey data from two independent samples ($n = 639$ and $n = 1,597$).

Individuals are asked to indicate their agreement or disagreement with prescriptive normative statements, written in the third person, along a five point Likert-type response format ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). For example, ‘It is wrong for a man to be seen in a gay bar’. Higher scores reflect endorsement of dominant representations of masculinity. Seventeen items are reverse scored to mitigate response bias.

Insert Table 1 about here.

MRNI (Levant, et al., 1992). The MRNI is a 57 item measure of masculinity ideology/ies. It belongs to the broader class of measure that assesses gender ideology/ies. Items were initially developed in order to operationalize seven theoretical norms of traditional masculinity, reflected in the subscales of Avoidance of Femininity (seven items), Rejection of Homosexuals (four items), Self-Reliance (seven items), Aggression (five items), Achievement/Status (seven items), Attitudes toward Sex (eight items), and Restrictive Emotionality (seven items). An additional norm and subscale of Nontraditional Attitudes toward Masculinity (12 items) was subsequently added. Individuals are asked to indicate their agreement or disagreement with descriptive normative statements, written in the third person, along a seven point Likert-type response format ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). For example, ‘If necessary a man should sacrifice personal relationships for career advancement’. Higher scores indicate greater endorsement of ideas and beliefs relating to traditional masculinity. The reverse is true for the Non-Traditional Attitudes subscale. Total scale and subscale scores are calculated by adding item scores for each scale and dividing by the respective number of items on each. The total scale score excludes item scores from the Nontraditional Attitudes subscale (Levant & Fischer, 1998).

Levant and Richmond (2007) summarize research having used the MRNI over the preceding 15 years. This offers some evidence in support of the Inventory’s use, particularly within the United States. Evidence exists of test-retest reliability over a 3-month period where $r = .65$ for men and $.72$ for women. The internal consistency of the total scale (i.e., α ranging between $.84$ and $.88$) and subscales (i.e., α ranging between $.52$ and $.83$) are also comparable across published studies. These make use of student samples (Russia and the United States; Levant, Cuthbert, et al., 2003; United States; Levant & Majors, 1997; China and the United States; Levant, Wu, & Fischer, 1996). Similar reliability coefficients appeared in this study (i.e., $\alpha = .92$ for the total scale and subscale coefficients ranged between $.46$ and

.83) (see Table 2 for a summary of scale mean, standard deviation and internal consistency values). The oft low reliability of subscales has resulted in these being excluded from analyses on occasion. Furthermore evidence in support of the Inventory's structural validity is lacking. Confirmatory factor analysis (CFA), for example, failed to replicate the seven theoretical norms that informed the development of subscales. This rather suggested three underlying factors (Levant, et al., 1992).

Insert Table 2 about here.

GCRS (O'Neil, et al., 1986). The GCRS is a 37 item measure. It seeks to operationalize the assumptions of gender role conflict theory (O'Neil, Good & Holmes, 1995). O'Neil and Good (1997, p. 11) argue that "(g)ender role conflict occurs when rigid, sexist, or restricted gender roles result in restriction, devaluation, or violations of others or self". The measure assesses men's gender role conflict in terms four specific patterns. These are described by the subscales of Success, Power and Competition (13 items), Restrictive Emotionality (10 items), Restricted Affectionate Behavior Between Men (eight items), and Conflict Between Work and Family Relations (six items). Individuals are asked to indicate their agreement or disagreement with descriptive normative statements, written in the first person, along a six point Likert-type response format ranging from 1 ("strongly disagree") to 6 ("strongly agree"). For example, 'I feel torn between my hectic work schedule and caring for my health'. Higher scores indicate greater levels of conflict. Total scale and subscale scores may be calculated by adding item scores for each scale and dividing by the respective number of items on each (O'Neil, 2008).

O'Neil (2008) summarizes research having used the GRCS over the preceding 25 years. This provides evidence in support of the instruments use, cross-nationally, but again

primarily within the United States. Its factor structure is replicated in studies assessing diverse populations and explains between 32% and 52% of variance. Likewise internal consistency of the total scale (i.e., α ranging between .70 and .89) and subscales (i.e., α ranging between .71 and .91) remains similar across these (e.g., Sweden; Bjerke & Skyllingstad, 2002; Korea; Kim, Hwong & Ryu, 2003; United States; O'Neil, et al., 1986). This is also the case for the current study (i.e., total scale $\alpha = .88$ and subscale reliability ranged between .77 and .86) (see Table 2 for a summary of scale mean, standard deviation and internal consistency values). Other evidence cited in support of the measure's use includes test-retest reliability over a 1-month period (i.e., r ranging between .72 and .86 across subscales) (O'Neil, 2008).

PAQ (Spence & Helmreich, 1978; Spence, et al., 1974). The PAQ (Spence, et al., 1974) is a 24 item measure originally proposed as a means to test masculinity and femininity. As a measure of gender orientation it seeks to establish people's gender, in the form of personality traits, through having them rate themselves in terms of stereotypical masculine or feminine characteristics (Thompson & Pleck, 1995). The PAQ includes Masculine (i.e., M, eight items, "Not at all independent – Very independent"), Masculine-Feminine (i.e., M-F, eight items, "Not at all aggressive – Very aggressive") and Feminine (i.e., F, eight items, "Not at all emotional – Very emotional") scales. Individuals are asked to indicate their similarity to dichotomous personality descriptors along a five point scale ranging from 0 ("A") to 4 ("E"). Higher scale scores indicate greater levels of either masculinity or femininity. Six items are reverse scored. Scale scores may be calculated through summing item scores for each. This study makes use of the eight item PAQ - Masculinity scale (PAQ-M) which is specifically meant to measure an individual's masculine personality traits.

There is mixed evidence in support of the PAQ's structural validity. Factor analyses have rendered between two and four factors, accounting for approximately 40% of variance,

depending upon whether all three scales have been included in analysis (Hill, Fekken & Bond, 2001). Helmreich, Spence and Wilhelm (United States; 1981), for example, demonstrated that M and F items loaded onto two orthogonal factors as expected. These have alternatively been described as assessing the traits of 'instrumentality' and 'expressiveness' (Hoffman, 2001) or 'agency' and 'communion' (Hill, et al., 2001). The internal consistency of these scales varies quite widely (i.e., M α ranging between .51 and .82 whereas F α ranging between .65 and .82) (Hill, Fekken & Bond, 2001). A $\alpha = .70$ emerged for this study (see Table 2 for a summary of scale mean, standard deviation and internal consistency values). Whilst the majority of studies providing measurement validity evidence make use of student samples, primarily from the United States (e.g., United States; Cota & Fekken; United States; Ward, Thorn, Clements, Dixon & Sanford, 2006), some have been conducted elsewhere (e.g., United Kingdom; McCreary & Steinberg, 1992; Spain; Fernández & Coello, 2010).

Procedure

A trained research assistant distributed paper-and-pencil questionnaires. Questionnaire completion took place during formal lectures or in the library. Instructions appeared in writing but were also emphasized by the assistant before questionnaire completion. It was stressed that the questionnaire should be completed alone and in silence. Participants were asked to ensure that they provided a single response to each question or statement. Standard ethical guidelines were adhered to. Participants were assured of their anonymity, confidentiality of data, and were informed of their right to withdraw from the study prior to and during its conduct. They were required to provide informed written consent. The procedure took between 20 and 30 minutes in each case and questionnaires were returned immediately after completion. The research assistant remained present at all times.

Participants were offered a briefing sheet after their involvement as well as a small monetary contribution toward out-of-pocket expenses.

Fatigue effects did not appear to have impacted upon the quality of data. Blasius and Thiessen (2012) suggest that the presence of these effects may be assessed through considering the reliability of items appearing toward the end of the questionnaire which, when such effects are present, tend to be lower than items appearing at the beginning. A useful measure of this was to compare the internal consistency of MRI total scale items appearing at the beginning of the questionnaire (Version 1, 3 and 5; $n = 159$; $\alpha = .88$) against those appearing at its end (Version 2, 4 and 6; $n = 160$; $\alpha = .89$). Internal consistency scores appeared comparable.

Results

Hypothesis 1

The hypothesis was supported. EFA confirmed construct multidimensionality. Interpretable simple structure was achieved and informed largely reliable total and subscale measures. Moreover subscales demonstrated significantly higher positive correlations with the total scale than with one another.

Data were suitable for EFA. An acceptable subject-to-variable (SVR) ratio of 6.65:1 (Kass & Tinsley, 1979) and a sample size > 300 (Comrey & Lee, 1992) were obtained. The factorability of the correlation matrix was determined through two established criteria. It contained a number of coefficients $\geq .45$ (Tabachnick & Fidell, 2007) and Bartlett's test of sphericity was significant [$\chi^2(4191.22)$; $df = 1128$; $p < .01$] (Bartlett, 1954). A Kaiser-Meyer-Olkin value of .84 also indicated sampling adequacy (Kaiser, 1974).

Principal components analysis (PCA) was adopted as the preferred method of analysis, as suggested by Fischer and Fontaine (2011), for exploratory purposes in cross-cultural research. Costello and Osborne (2005) note the often heated debate concerning the

relative merit of PCA as opposed to factor analysis. In this study, initial exploration of data through both supports the argument that, given adequate sample size and quality of data, there is often little difference between PCA and FA (i.e., principal axis factoring) in obtained results (DeVellis, 2012; Tabachnick & Fidell, 2007).

Although there are theoretical reasons to suppose that underlying factors may be related in this study, oblique rotation demonstrated weak correlations (Cohen, 1992) between most factors ($r = .02; .15; .31$) Penhazur and Schmelkin (1991) argue that in such cases it is most reasonable to make use of orthogonal rotation.

Fifteen factors emerged through applying an eigenvalue extraction criterion of ≥ 1.00 (Kaiser, 1960). Yet this solution proved difficult to interpret prior to and following rotation. Items failed to load substantively onto a number of factors at $\geq .35$ (Hair, Tatham, Anderson & Black, 1998). Parallel analysis alternatively suggested the worth of extracting five factors (Horn, 1965). But this this solution did not achieve simple structure prior to and following rotation. Finally consideration of the scree plot indicated extracting three factors above the point of inflexion (Cattell, 1966).

The resultant factor solution achieved simple structure after orthogonal varimax rotation and was interpretable (see Table 3). This accounted for 28.29% of total variance (i.e., 17.05%, 6.62% and 4.61% for successive factors). Twelve items failed to achieve a minimum factor loading of $\geq .35$ and were removed. A further four items were also removed as they demonstrated poor conceptual fit and relatively low loadings compared to other items on relevant factors. Their removal did not impact adversely upon internal consistency which was further enhanced through the removal of three items with an item-total correlation of $< .2$ with other items on the corresponding factor (DeVellis, 2012). Remaining items informed the MRI total scale consisting of 29 items ($\alpha = .89; M = 3.13; SD = .50; \text{range} = 29 \text{ to } 145$) and

three subscale measures. Table 2 summarises the mean, standard deviation and internal consistency of each.

Insert Table 3 about here.

The first subscale consists of 11 items ($\alpha = .85$, $M = 2.65$, $SD = .74$, range = 11 to 55) loading substantively on Factor 1. These may be described as reflecting the dominant Representation of Undesirables Others. Endorsement of these items contrasts the well-documented co-construction of effeminacy and homosexuality against ‘real’ masculinity (Gough & Edwards 1998; Luyt, 2012). That is to say, to be a real man one should neither engage in practices associated with homosexuality (e.g., item 24, ‘It is wrong for a man to be seen in a gay bar’) nor supposedly related effeminate practices (e.g., item 18, ‘Men who cry out-loud in public are weak’).

The second subscale contains 14 items ($\alpha = .83$, $M = 3.76$, $SD = .53$, range = 14 to 70) loading substantively on Factor 2. These capture what may be described as the dominant Representation of Responsibility. Items underline the importance of men’s dependability (e.g., item 10, ‘A true friend is someone who would fight by a man’s side no matter what’) and success (e.g., item 30, ‘It is important for a man to be successful in his work’).

The third subscale comprises of 4 items ($\alpha = .56$, $M = 2.29$, $SD = .63$, range = 4 to 20) loading substantively on Factor 3. These reveal the importance of dominance (e.g., item 38, ‘A man’s decision should not be questioned’) and toughness [e.g., item 48, ‘A man should (not) tell others when he is feeling depressed’] in defining ‘appropriate’ masculine practice. They contribute toward the dominant Representation of Control.

Comparison between total-subscale correlations and subscale correlations suggest that subscales measure different dimensions of a common construct. That is to say, subscales

exhibited significantly higher positive correlations with the total scale (i.e., r ranging between .60 and .87) than with one another (i.e., r ranging between .34 and .46) (see Table 4). The significance of these differences was confirmed through Fisher's r -to- z transformation (z ranging between -2.46 and -12.31) (see Table 5).

Insert Table 4 about here.

Insert Table 5 about here.

Hypothesis 2

The hypothesis was supported. MRI scale scores demonstrated significant positive and negative correlation with those of the MRNI (see Table 6). Total scale scores correlated ($r = .87, p < .01$) as did conceptually related subscale scores. For example, as might be expected, the MRI Undesirable Others and MRNI Avoidance of Femininity ($r = .75, p < .01$) as well as the Rejection of Homosexual subscales ($r = .75, p < .01$); the MRI Responsibility and MRNI Self-Reliance ($r = .57, p < .01$) subscales; and the MRI Control and MRNI Restrictive Emotionality ($r = .67, p < .01$) subscales were most strongly positively correlated with one another. The MRNI Nontraditional Attitudes towards Masculinity subscale was, on the other hand, meaningfully negatively correlated with MRI scales (i.e., r ranging between -.43 and -.54, $p < .01$). These results provide evidence in support of MRI scale scores moderate to strong (Dancey & Reidy, 2004) convergent validity with those of the MRNI. As measures of gender (re)presentation and gender ideology/ies, they appear most closely related.

Insert Table 6 about here.

Hypothesis 3

The hypothesis was supported. MRI scale scores demonstrated significant positive correlation with those of the GRCS (see Table 6). Total scale scores correlated ($r = .45, p < .01$) as did conceptually related subscale scores. For example, as might be anticipated, the MRI Undesirable Others and GRCS Restrictive Affectionate Behavior Between Men subscales ($r = .49, p < .01$); the MRI Responsibility and GRCS Success/Power/Competition ($r = .43, p < .01$) as well as Conflict Between Work and Family Relations subscales ($r = .23, p < .05$); and the MRI Control and GRCS Restrictive Emotionality ($r = .37, p < .01$) subscales were most strongly positively correlated with one another. These results provide evidence in support of MRI scale scores low to moderate (Dancey & Reidy, 2004) convergent validity with those of the GRCS. As measures of gender (re)presentation and gender role conflict, whilst they appear related, they do not demonstrate as strong convergence as compared to measures of gender ideology/ies.

Hypothesis 4

The hypothesis was largely supported. MRI scale scores demonstrated nonsignificant correlations (r ranging between -0.10 and $.06$, NS), in all but one case, with the PAQ-M scale score (see Table 6). That is to say, a low, but significant positive correlation ($r = .20, p < .05$) appeared between the MRI Responsibility subscale and the PAQ-M scale. These results, for the most part, provide evidence in support of MRI scale scores divergent validity with that of the PAQ-M. As a measure of gender orientation, the PAQ demonstrates the weakest relationship to the MRI, as compared to measures of gender ideology/ies and gender role conflict.

Discussion

This study contributes toward research concerning the development of the MRI and associated theorizing. It provides measurement validity evidence in support of the MRI's use among English speakers in South Africa; albeit somewhat limited through the use of a student sample. This includes information concerning its structural, convergent and divergent validity. Three interpretable factors emerged through EFA, thereby confirming construct multidimensionality. These accounted for 28.29% of total variance. This is in close approximation to other similar studies making use of the MRNS (28%; Thompson & Pleck, 1986), the MANI-II (31.44%; Luyt, 2005), and the GRCS (36%; O'Neil, et al., 1986). Relatively low total variance such as these are more common among factors that may be thought of as more abstract, or conceptually distant, from their constituent items. This may arguably be the case in measures assessing social norms. Whilst an acceptable SVR ratio and a sample size were obtained for the purpose of EFA, it is likely that a larger sample, approximating the recommended SVR ratio of 1:10 (Nunnally, 1978), would be beneficial in increasing explained total variance. Cross-loading items are particularly likely to appear in the initial stage of measurement construction (Osborne & Costello, 2004), where the SVR ratio or sample size is moderate, this resulting in lower total variance accounted for.

Total and subscale measures were subsequently developed. Subscale measures assess the endorsement of dominant gender representations, including Othering (Anti-Effeminacy and Homo-Negativity), Responsibility (Dependability and Success) and Control (Dominance and Toughness). The subscale of Responsibility reflects the added emphasis afforded this theoretical dimension following cross-cultural content validation and subsequent revision of the MANI-II. Both the Othering and Responsibility subscales represent novel dimensions in masculinity-related measurement literature. As predicted, subscale scores showed significantly higher positive correlation with the total scale score than with one another, suggesting that they measure different dimensions of a common construct.

The internal consistency of these subscales was high in all but one case. That is to say, the subscale of Control demonstrated a low Cronbach's alpha value ($\alpha = .56$). This may be due to its brevity (Cronbach, 1951), including only four items, as well as a moderate sample size of approximately 300. Charter (1999), for example, suggests that a minimum sample size of 400 is required for a more accurate estimate of the population coefficient alpha. Apart from its potential effect in reducing the alpha value, its brevity relative to the Othering and Responsibility subscales, should not be cause for concern. Its development was empirically informed and disparity in subscale length is a feature of other gender-related measures such as the CMNI (Mahalik, et al., 2003). It may also be argued that the low alpha value reflects item heterogeneity, or more specifically, that dominance and toughness items do not sufficiently measure the same overarching construct of Control. The same might be argued in the case of Reardon and Govender's (2013) use of the MANI-II where a similar Control subscale demonstrated a low Cronbach's alpha value of .55; although again through the use of a debatably small sample ($n = 157$). Yet interpretation of the related factor as well as consideration of item inter-correlations, which are all of a similar size, does not support this argument. It is also worthwhile to contemplate the counter-argument that some degree of item heterogeneity is worthwhile in order to ensure that each is measuring a slightly different aspect of the overarching construct. That is to say, item homogeneity may on occasion reflect item redundancy rather than internal consistency (Boyle, 1991). The Control subscale's low alpha value indicates that any related results should be treated with some caution. Future research might examine whether this value increases in the case of a larger sample size and whether items might suggest additional overarching constructs.

Convergent and divergent validity evidence was obtained through a largely anticipated pattern of relationships between the MRI, MRNI, GRCS and PAQ-M scale scores. Four of the MRNI subscales exhibited low internal consistency (i.e., Rejection of

Homosexuals, $\alpha = .56$; Self-reliance, $\alpha = .46$; Aggression, $\alpha = .58$; Nontraditional Attitudes toward Masculinity, $\alpha = .56$). This is a feature, as noted above, in other studies making use of the MRNI subscales. This may reflect the fact that the measure lacks evidence in support of its structural validity. In this respect it is noteworthy that the GRCS and PAQ-M scales demonstrated moderate to high internal consistency in this study. It may also be a sign of the fact that it was not developed specifically for use in South Africa. This serves to underline the importance of developing gender-related measures for which validity evidence exists in support of their use among specific populations. It also highlights that difficulty of undertaking measurement validation in contexts such as South Africa where few measures of this kind exist. Correlations with these subscales should therefore be cautiously interpreted.

Despite the low internal consistency of some MRNI subscales, the MRI and the MRNI total scale scores correlated strongly ($r = .87$) as hypothesized. The strength of this correlation may imply that the two measures assess the same construct and, such as, development of the MRI is unnecessary. Yet, even though the MRNI is described as a measure of gender ideology/ies whilst the MRI is described as a measure of gender (re)presentation, it is important to remember that they share many conceptual assumptions. This is reflected, for example, in the similar wording of items. But the two measures do differ conceptually in their emphasis on analyzing data at the group or social level as opposed to the individual level and have been developed within different cultural contexts. This different analytic emphasis, and the cultural sensitivity of measurement, will not necessarily lessen the statistical strength of relationship between these measures.

Secondly, this paper adopts an 'indigenous' (van de Vijver & Tanzer, 2004), 'particularizing' (Adcock & Collier, 2001) or 'emic' (Benet-Martínez, 2009) approach to measurement. This assumes that a construct may not be understood in the same way by socio-cultural groups due to basic differences in understanding. Rather than either assuming

measurement equivalence or seeking to obtain limited evidence concerning only the structural validity of existing measures such as the MRNI, as is most often done, effort begins by evaluating content validity evidence in support of a measure's use. This should ideally involve a mixed method (QUAN + QUAL) approach that inductively explores how these groups understand gender. Measurement equivalence is neither considered desirable nor necessarily possible at the outset. Measures should rather reflect complexity in socio-cultural understanding of gender through operationalizing culture specific constructs where necessary. Although this potentially complicates or disallows cross-cultural comparison, and sets hurdles for related interpretation, richer findings may result (Luyt 2012, 2013). The fact that insufficient evidence exists in support of the MANI-II's cross-cultural measurement validity in South Africa was a primary motivation for this paper.

An unexpected significant positive correlation ($r = .20, p < .05$) also appeared between the MRI Responsibility subscale and the PAQ-M scale. This might be explained by their similar emphasis on concepts such as dependability, independence and success. This correlation was nonetheless low. The theoretical relationship between measures of gender (re)presentation, gender ideology/ies, gender role conflict and gender orientation was supported, as predicted. That is to say, the measure of gender ideology/ies (i.e., the MRNI) was most closely related to the MRI as a measure of gender (re)presentation, followed by the measure of gender role conflict (i.e., the GRCS). The PAQ-M, a measure of gender orientation, demonstrated the weakest relationship.

The MRI may be described as a novel multidimensional measure of gender (re)presentation. It is designed to index socio-cultural group endorsement of dominant gender representations. Crucially, in being informed by social constructionism, it seeks to analyze data at the group or social level. It also adopts an emic approach to measurement where cross-cultural measurement equivalence is neither considered desirable nor necessarily

possible at the outset. The MRI was developed through the revision of the MANI-II which was guided by evidence relating to its cross-cultural validity (Luyt, 2013/2015). Whilst content and construct validity evidence exists in support of the MRI's use among English speaking South African's, further construct validity evidence should be obtained in support of its use among other language groups in the country such as Afrikaans and Xhosa speakers. In doing so, future studies should seek to reflect complexity in socio-cultural understanding through operationalizing culture specific constructs where necessary. The generalizability of findings within socio-cultural groups should ideally be enhanced in future through the use of more representative samples extending beyond student populations. Online survey panels increasingly offer an effective way of achieving this in highly demographically diverse countries, within and across geographical region, such as South Africa.

It is suggested that a number of steps might usefully be adopted in a measure's cross-cultural construct validation. Crucially, this should only take place once evidence in support of the measure's cross-cultural content validity has been obtained, in order to ensure that the measure adequately captures cultural understanding of the construct. Firstly, RFA (Ben-Porath, 1990) should be applied in order to determine whether similar factors emerge from data obtained from the new socio-cultural group as compared to analysis of data from the original socio-cultural group. This method offers a less stringent means of assessing factorial invariance than CFA procedures. In doing so, it facilitates exploration concerning the degree to which measures and their associated scales are similar, and hence the extent to which cross-cultural comparisons can be made.

The same EFA procedures for estimating communalities and rotation, which were used in the original analysis, should be applied. The number of factors extracted should also be restricted to the number emerging from the original data. If simple structure with similar meaning does not emerge, then the analysis should be repeated based upon conventional

procedures of factor extraction. Where resultant factors may be interpreted to hold the same meaning, coefficients of congruence, such as Tucker's phi (Tucker, 1951), should be calculated in order to provide a formal test of factor similarity. Only in instances where coefficients indicate congruence, should more stringent CFA procedures be applied in order to determine factorial invariance, for example, configural, metric and scalar invariance (Milfont, & Fischer, 2010).

This suggested procedure will indicate whether measures and their associated scales hold no conceptual similarity, conceptual similarity but not measurement equivalence, or varying forms of measurement equivalence cross-culturally. It is plausible, therefore, that a measure might only be partially cross-culturally congruent where it variably contains culture specific subscales with no similarity, those with conceptual similarity, and others that hold varying forms of measurement equivalence. This would occasion a set of cross-cultural comparisons which would be more complex, less neat, and far more difficult to interpret. Yet these would be both more meaningful and richer for the effort.

Such results would be made even more meaningful through obtaining criterion-related evidence in support of total and subscale measures. Concurrent validity evidence is particularly useful in indicating a relationship between measures, especially subscale measures, and relevant outcome variables that are assessed at the same time. That is to say, given the social constructionist assumption that people adopt inconsistent or contradictory gender practices – including in the form of expressed attitudes, beliefs and ideas – gathering evidence of predictive validity is less meaningful. This seeks to establish a relationship between measures and outcome variables over a longer term which is not assumed.

Measurement validity evidence exists in support of the MRI's use, as the only available measure of masculinity (re)presentation, among English speakers in South Africa. Its applied utility lies in our being able to index socio-cultural group endorsement of

dominant gender representations. In identifying dominant gender representations that (re)produce unequal gender relations, resultant practical interventions aimed at promoting more equitable gender relations, will therefore be aimed at social rather individual processes – gender is understood as a social practice. This differs from understanding gender as a set of individual attributes, characteristics or traits, which would recommend interventions at, for example, an individual therapeutic level (e.g., clinical or counselling practice). Willig (1999, p. 15), for instance, underlines the utility of discourse analysis in informing the development of practical interventions. In doing so, “it seeks to expose the ways in which language is used ideologically to maintain unequal power relations in society, but it also aspires to bring about positive change in institutional and social practices”. Lamerichs and te Molder’s (2011) Discursive Action Method (DAM) offers a specific example of a workable practical intervention. This seeks to shape social practices through making individuals critically aware of the way in which they talk about issues within social groups. It is in the social (re)production of gender that, through collectively (re)imagining gender possibilities, individual change is possible.

The immediate challenge is now to determine what evidence exists in support of the MRI’s cross-cultural construct validity. This will allow us to understand how normative or dominant gender representations are understood by socio-cultural groups. It will also provide a means with which to explore their variable endorsement across and within socio-cultural groups; the potential functions such variations serve; and the subject positions that are made possible as a result.

References

- Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, *95*, 529-546.
- Baber, K.M., & Tucker, C. J. (2006). The Social Roles Questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles*, *54*, 459-467. doi: 10.1007/s11199-006-9018-y
- Blasius, J., & Thiessen, V. (2012). *Assessing the Quality of Survey Data*. London: Sage Publications.
- Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society Series B*, *16*, 296-298.
- Beall, A. E. (1993). A social constructionist view of gender. In A. E. Beall & R. J. Sternberg (Eds.), *The Psychology of Gender* (pp. 127-147). New York, NY: Guilford Press.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Clinical and Consulting Psychology*, *42*, 155-162. doi: 10.1037/h0036215
- Benet-Martínez, V. (2009). Cross-cultural personality research: Conceptual or methodological issues. In R. W. Robins, R. C. Fraley & R.F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 170-189). London: Guilford Press.
- Ben-Porath, Y. S. (1990). Cross-cultural assessment of personality: The case for replicatory factor analysis. In N. Butcher & C. D. Spielberger (Eds.), *Advances in Personality Assessment* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum.
- Bjerke, A., & Skyllingstad, Y. E. (2002). Gender Role Conflict Scale in Sweden. *Psykologexamensuppsata*, *4*, 1-29.
- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometrics scales? *Personality and Individual Differences*, *12*, 291-294.

- Brannon, R., & Juni, S. (1984). A scale for measuring attitudes about masculinity. *Psychological Documents*, 14, 6-7.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. doi: 10.1191/1478088706qp063oa
- Brislin, R. W. (2000). Back-translation. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (pp. 359-360). Washington, DC: American Psychological Association.
- Burr, V. (1998). *Gender and Social Psychology*. Hove: Routledge.
- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21, 559-566.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ: Erlbaum.
- Connell, R. W. (1987). *Gender and Power: Society, the Person and Sexual Politics*. Cambridge: Polity Press.
- Connell, R. W. (1990). An iron man: The body and some contradictions of hegemonic masculinity. In M. A. Messner & D. F. Sabo (Eds.), *Sport, Men, and the Gender Order: Critical Feminist Perspectives* (pp. 83-95). Champaign, IL: Human Kinetics Books.
- Connell, R. W. (1993). The big picture: Masculinities in recent world history. *Theory and Society*, 22, 597-623. doi: 10.1007/BF00993538
- Connell, R. W. (1995). *Masculinities*. Cambridge: Polity Press.
- Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin*, 80, 389-407. doi: 10.1037/h0035334
- Coopersmith, S. (1967). *The Antecedents of Self-Esteem*. San Francisco: Freeman.

- Cota, A. A., & Fekken, G. C. (1988). Dimensionality of the Personal Attributes Questionnaire: An empirical replication. *Journal of Social Behavior and Personality*, 3, 135-140.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dancey, C., & Reidy, J. (2004). *Statistics without Maths for Psychology: Using SPSS for Windows*. London: Prentice Hall.
- Davis, S. N., & Greenstein, T. N. (2009). Gender ideology: Components, predictors, and consequences. *Annual Review of Sociology*, 35, 87-105. doi: 10.1146/annurev-soc-070308-115920
- DeVellis, R. F. (2012). *Scale Development: Theory and Applications*. Thousand Oaks, CA: Sage Publications.
- Edley, N. (2001). Unravelling social constructionism. *Theory and Psychology*, 11, 433-441. doi: 10.1177/0959354301113008
- Eisler, R. M., & Skidmore, J. R. (1987). Masculine gender role stress: Scale development and component factors in the appraisal of stressful situations. *Behavior Modification*, 11, 123-136. doi: 10.1177/01454455870112001
- Fernández, J., & Coello, M. T. (2010). Do the BSRI and PAQ Really Measure Masculinity and Femininity? *Spanish Journal of Psychology*, 13, 1000-1009.
- Fischer, R., & Fontaine, R.J. (2011). Methods for investigating structural equivalence. In D. Matsumoto & F.J.R. van de Vijver (Eds.), *Cross-Cultural Research Methods in Psychology* (pp. 179-215). Cambridge: Cambridge University Press.
- Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American*

Psychologist, 40, 266-275. doi: 10.1037/0003-066X.40.3.266

Gibbons, J. L., Hamby, B. A., & Dennis, W. D. (1997). Researching gender-role ideologies internationally and cross-culturally. *Psychology of Women Quarterly*, 21, 151-170.

doi: 10.1111/j.1471-6402.1997.tb00106.x

Gough, B., & Edwards, G. (1998). The beer talking: Four lads, a carry out and the reproduction of masculinities. *Sociological Review*, 46, 409-435. doi: 10.1111/1467-954X.00125

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate Data Analysis*. Prentice-Hall: London.

Hearn, J. (2010). Reflecting on men and social policy: Contemporary critical debates and implications for social policy. *Critical Social Policy*, 30, 165-188. doi:

10.1177/0261018309358288

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. (1981). A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles*, 7, 1097-1108. doi:

10.1007/BF00287587

Hill, S. A., Fekken, G. C., & Bond, S. L. (2000). Factor structure integrity of the Personal Attributes Questionnaire: An English-French comparison. *Canadian Journal of Behavioural Science*, 32, 234-242. doi: 10.1037/h0087120

Behavioural Science, 32, 234-242. doi: 10.1037/h0087120

Hoffman, R. M. (2001). The measurement of masculinity and femininity: Historical perspective and implications for counselling. *Journal of Counseling and Development*, 79, 472-485.

Hogan, T. P., & Agnello, J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational and Psychological Measurement*, 64, 802-812.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

- Joint Committee on Standards for Educational Evaluation (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kass, R. A., & Tinsley, H. E. A. (1979). Factor analysis. *Journal of Leisure Research*, 11, 120-138.
- Kim, J., Hwong, M., & Ryu, J. (2003). The cultural validation of the Korean version of the GRCS (K-GRCS). *Asian Journal of Education*, 4, 25-41.
- Kottak, C. (2012). *Mirror for Humanity: A Concise Introduction to Cultural Anthropology*. New York, NY: McGraw-Hill.
- Kroska, A. (2000). Conceptualising and measuring gender ideology as an identity. *Gender and Society*, 14, 368-394. doi: 10.1177/089124300014003002
- Lamerichs, J., & te Molder, H. (2011). Reflecting on your own talk: The Discursive Action Method at work. In C. Antaki (Ed.), *Applied Conversation Analysis* (pp. 184-206). Basingstoke: Palgrave Macmillan.
- Levant, R. F., & Fischer, J. (1998). The Male Role Norms Inventory. In M. Davis, W. H. Yarber, R. Bauserman, G. Schree & S. L. Davis (Eds.), *Sexuality-Related Measures: A Compendium* (pp. 469-472). Thousand Oaks, CA: Sage Publications.
- Levant, R. F., & Majors, R. G. (1997). Masculinity ideology among African American and European American college women and men. *Journal of Gender, Culture and Health*, 2, 33-43.
- Levant, R. F., & Richmond, K. (2007). A review of research on masculinity ideologies using the Male Role Norms Inventory. *Journal of Men's Studies*, 15, 130-146. doi: 10.3149/jms.1502.130

- Levant, R. F., Cuthbert, A., Richmond, K., Sellers, A., Matveev, A., Mitina, O., ...
- Heesacker, M. (2003). Masculinity ideology among Russian and U.S. young men and women and its relationship to unhealthy lifestyle habits among young Russian men. *Psychology of Men and Masculinity, 4*, 26-36. doi: 10.1037/1524-9220.4.1.26
- Levant, R. F., Hirsch, L., Celentano, E., Cozza, T., Hill, S., MacEachern, M., ... Schnedeker, J. (1992). The male role: An investigation of contemporary norms. *Journal of Mental Health Counseling, 14*, 325-337.
- Levant, R. F., Rankin, T. J., Williams, C. M., Hasan, N. T., & Smalley K. B. (2010). Evaluation of the factor structure and construct validity of scores on the Male Role Norms Inventory - Revised (MRNI-R). *Psychology of Men & Masculinity, 11*, 25-37. doi: 10.1037/a0017637
- Levant, R. F., Richmond, K., Cook, S., House, A. T., & Aupont, M. (2007). The Femininity Ideology Scale: Factor structure, reliability, convergent and discriminant validity, and social contextual variation. *Sex Roles, 57*, 373-383.
- Levant, R. F., Richmond, K., Majors, R. G., Inclan, J. E., Rosello, J. M., Rowan, G., & Sellers, A. (2003). A multicultural investigation of masculinity ideology and alexithymia. *Psychology of Men and Masculinity, 4*, 91-99.
- Levant, R. F., Smalley, K. B., Aupont, M., House, A., Richmond, K., & Noronha, D. (2007). Validation of the Male Role Norms Inventory - Revised. *Journal of Men's Studies, 15*, 85-100.
- Levant, R. F., Wu, R., & Fischer, J. (1996). Masculinity ideology: A comparison between U.S. and Chinese young men and women. *Journal of Gender, Culture and Health, 1*, 207-220.
- Luyt, R. (2013/2015). Beyond traditional understanding of gender measurement: The gender (re)presentation approach. *Journal of Gender Studies, 24*, 207-226.

- Luyt, R. (2012). A framework for mixing qualitative and quantitative methods in measurement development, validation and revision: A case study. *Journal of Mixed Method Research, 6*, 294-316.
- Luyt, R. (2005). The Male Attitude Norms Inventory-II: A measure of masculinity ideology in South Africa. *Men and Masculinities, 8*, 208-229.
- Luyt, R. (2003). Rhetorical representations of masculinities in South Africa: Moving towards a material-discursive understanding of men. *Journal of Community and Applied Social Psychology, 13*, 46-69.
- Luyt, R., & Foster, D. (2001). Hegemonic masculine conceptualisation in gang culture. *South African Journal of Psychology, 31*, 1-11.
- Marecek, J., Crawford, M., & Popp, D. (2004). On the construction of gender, sex, and sexualities. In A. H. Eagly, A. E. Beall & R. J. Sternberg (Eds.), *The Psychology of Gender* (pp. 192-216). London: Guildford Press.
- Morawski, J. G. (1987). The troubled quest for masculinity, femininity, and androgyny. In P. Shaver & C. Hendrick (Eds.), *Sex and Gender* (pp. 44-69). Newbury Park, CA: Sage Publications.
- Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P. J., Gottfried, M., & Freitas, G. (2003). Development of the Conformity to Masculine Norms Inventory. *Psychology of Men and Masculinity, 4*, 3-25.
- McCreary, D. R., Newcomb, M. D., & Sadava, S. W. (1998). Dimensions of the male gender role: A confirmatory analysis in men and women. *Sex Roles, 39*, 81-95.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*, 111-121.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGrawHill.

- O'Neil, J. M. (2008). Summarizing 25 years of research on men's gender role conflict using the Gender Role Conflict Scale: New research paradigms and clinical implications. *Counseling Psychologist, 36*, 358-445.
- O'Neil, J. M., & Good, G. E. (1997). Men's gender role conflict: Personal reflections and overview of recent research (1994-1997). *Society for the Psychological Study of Men and Masculinity (SPSMM) Bulletin, 3*, 10-15.
- O'Neil, J. M., Good, G. E., & Holmes, S. (1995). Fifteen years of theory and research on men's gender role conflict: New paradigms for empirical research. In R. F. Levant & W. S. Pollack (Eds.), *A New Psychology of Men* (pp. 164-206). New York, NY: Basic Books.
- O'Neil, J. M., Helms, B. J., Gable, R. K., David, L., & Wrightsman, L. S. (1986). Gender Role Conflict Scale: College men's fear of femininity. *Sex Roles, 14*, 335-350.
- Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation, 9*, 1-9.
- Pedhazur, E., & Schmelkin, L. (1991). *Measurement, Design and Analysis: An Integrated Approach*. Hillsdale, N.J.: Erlbaum.
- Pleck, J. H. (1976). The male sex role: Definitions, problems, and sources of change. *Journal of Social Issues, 32*, 155-164.
- Pleck, J. H. (1987). The theory of male sex-role identity: Its rise and fall, 1936 to the present. In H. Brod (Ed.), *The Making of Masculinities* (pp. 21-38). Boston, MA: Allen and Unwin.
- Pleck, J. H. (1995). The gender role strain paradigm: An update. In R. F. Levant & W. S. Pollack (Eds.), *A New Psychology of Men* (pp. 11-32). New York, NY: Basic Books.
- Reardon, C. A., & Govender, K. (2013). Masculinities, cultural worldviews and risk perceptions among South African adolescent learners. *Journal of Risk Research, 16*, 753-770. doi:10.1080/13669877.2012.73782

- Smiler, A. P. (2004). Thirty years after the discovery of gender: Psychological concepts and measures of masculinity. *Sex Roles, 50*, 15-26. doi: 10.1023/B:SERS.0000011069.02279.4c
- Smiler, A. P., & Epstein, M. (2010). Measuring gender: options and issues. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of Gender Research in Psychology: Vol. 1. Gender Research in General and Experimental Psychology* (pp. 133-157). New York, NY: Springer.
- Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and Femininity: Their Psychological Dimensions, Correlates, and Antecedents*. Austin, TX: University of Texas Press.
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity-femininity. *Catalog of Selected Documents in Psychology, 4*, 43-44.
- Statistics South Africa (2013). *Mid-year population estimates 2013: Statistical release* (SSA Publication No. P0302). Retrieved from <http://www.statssa.gov.za/publications/P0302/P03022013.pdf>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston, MA: Allyn & Bacon.
- Thompson, E. H., & Bennett, K. M. (2015). Measurement of masculinity ideologies: A (critical) review. *Psychology of Men & Masculinity, 16*, 115-133.
- Thompson, E. H., & Pleck, J. H. (1995). Masculinity ideologies: A review of research instrumentation on men and masculinities. In R. F. Levant & W. S. Pollack (Eds.), *A New Psychology of Men* (pp. 129-163). New York, NY: Basic Books.
- Thompson, E. H., & Pleck, J. H. (1986). The structure of the male role norms. *American Behavioral Scientist, 29*, 531-543.

- Thompson, E. H., Pleck, J. H., & Ferrera, D. L. (1992). Men and masculinities: Scales for masculinity ideology and masculinity-related constructs. *Sex Roles, 27*, 573-607. doi: 10.1007/BF02651094
- Tucker, L. R. (1951). *A Method for Synthesis of Factor Analysis Studies* (No. 984). Washington, DC: Department of the Army.
- van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée, 54*, 119-135. doi:10.1016/j.erap.2003.12.004
- Ward, L. C. Thorn, B. E., Clements, K. L., Dixon, K. E., & Sanford, S. D. (2006). Measurement of agency, communion, and emotional vulnerability with the Personal Attributes Questionnaire. *Journal of Personality Assessment, 86*, 206-216. doi:10.1207/s15327752jpa8602_10
- West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender and Society, 1*, 125-151. doi: 10.1177/0891243287001002002
- Willig, C. (1999). Introduction: Making a Difference. In C. Willig (Ed.), *Applied Discourse Analysis* (pp. 1-21). Buckingham: Open University Press.

Figure and Tables

Table 1

The Systematized Concept of the Masculinities Attitude Norms Inventory-III Including Its Six Theoretical Dimensions and Their Underlying Concepts

Dimensions					
Sexuality	Toughness	Independence	Status	Responsibility	Homophobia
Underlying concepts					
Sexual objectification	Emotional containment	Assertive behavior	Behavior management	Duty	Homophobic ostracism
Sexual control	Emotional denial	Achievement management	Career management	Dependability	Homophobic violence
Sexual self-appraisal	Self-containment	Self-actualization	Resource management	Self-sacrifice	Anti-homoerotic behavior
Sexual other-appraisal	Physical tenacity	Interpersonal dominance	Power management	Accountability	Homophobic avoidance

Note. Adapted from Luyt (2012, p. 301).

Table 2

Means, Standard Deviations, and Cronbach's Alpha Coefficients for the MRI, CRGS, MRNI and PAQ-M Scale Scores

Scales	M	SD	α
MRI Total ^a	3.13	.50	.89
Othering (Anti-Effeminacy and Homo-Negativity)	2.65	.74	.85
Responsibility (Dependability and Success)	3.76	.53	.83
Control (Dominance and Toughness)	2.29	.63	.56
GRCS Total ^b	4.14	.60	.88
Success, Power and Competition	3.40	.86	.86
Restrictive Emotionality	3.44	.97	.85
Restricted Affectionate Behavior Between Men	3.76	1.04	.82
Conflict Between Work and Family Relations	3.76	.95	.77
MRNI Total ^c	3.88	.74	.92
Avoidance of Femininity	3.99	1.16	.83
Rejection of Homosexuals	3.65	1.21	.56
Self-Reliance	4.78	.67	.46
Aggression	4.89	.81	.58
Achievement/Status	3.69	.92	.66
Attitudes toward Sex	3.16	.92	.77
Restrictive Emotionality	3.38	1.03	.77
Nontraditional Attitudes toward Masculinity	4.22	.67	.56
PAQ-M ^d	21.96	4.53	.70

Note. ^a*n* = 319 after listwise deletion of missing data. ^b*n* = 102 after listwise deletion of missing data. ^c*n* = 105 after listwise deletion of missing data. ^d*n* = 109 after listwise deletion of missing data.

Table 3

The Masculinities Attitude Norms Inventory-III Factors and Item Loadings

Factors and Items		Loading
Factor 1: "The Representation of Undesirable Others" (Anti-Effeminacy and Homo-Negativity)		
1	A man should prefer sports to needlework.	.48
5	To be a man you need to be physically tough. ^b	.37
6	Being called a 'faggot' is one of the worst insults to a man.	.37
8	A father should not be embarrassed if he finds out that his son is gay. ^a	.70
12	Men should not be allowed to sleep intimately in the same bed together.	.66
18	Men who cry out-loud in public are weak.	.52
19	Men who stay at home to clean and look after the children should be proud of what they do. ^a	.45
24	It is wrong for a man to be seen in a gay bar.	.77
31	A man should not feel embarrassed that his best friend is gay. ^a	.77
34	Gay men should be beaten-up.	.66
37	Men should be able to kiss each other passionately without feeling ashamed. ^a	.72
39	It is a man's task to ask someone on a first date. ^b	.37
47	Gay men are not suited to many jobs.	.67
Factor 2: "The Representation of Responsibility" (Dependability and Success)		
3	A man should be able to provide for his family.	.43
9	Men should appear confident in difficult situations.	.56
10	A true friend is someone who would fight by a man's side no matter what.	.55
13	Men should do work that earns them respect.	.53
14	A successful man should be able to live a comfortable life.	.53
15	A man deserves the respect of his family.	.56
20	Men should feel embarrassed if they are unable to get an erection with a new sexual partner.	.38
25	A man should take the lead when something needs to be done.	.56
28	It is important for men to be good in bed.	.52
29	Men should be determined to do well.	.56
30	It is important for a man to be successful in his work.	.69
35	Men should be able to remain focussed even in difficult situations.	.53
36	A man should back his friends up no matter what.	.49
41	Men should aim to have the respect and admiration of others.	.50
Factor 3: "The Representation of Control" (Dominance and Toughness)		
2	If a man hurts himself he should try not to let others see he is in pain.	.36
7	A man should take a break from his responsibilities to be with friends. ^{ab}	-.39
22	If a man is frightened he should try and not to let others see it.	.36
26	A man need not plan well in advance for the future. ^{ab}	-.36
38	A man's decision should not be questioned.	.42
43	It is wrong for men to call anyone a 'chick'. ^{acd}	.44
45	Men should be careful not to take unnecessary risks. ^{ac}	.36
46	It is not necessary for a man to love his sexual partner. ^c	.46
48	A man should tell others when he is feeling depressed. ^a	.48

Note. *n* = 319 after listwise deletion of missing data. Items appearing in bold removed from scale. ^aItem reverse-scored. ^bItem removed from scale as it lacked conceptual fit and demonstrated relatively low loadings compared to others on the factor. ^cItem removed from scale as its item-total correlation was < 2. ^d'Chick' is South African slang used to refer to a girl or woman, often considered attractive, yet may be perceived as demeaning,

disparaging, insulting or offensive in connoting sexual objectification or infantilizing. It is similar to the slang use of the word 'baby'.

Table 4

Correlation Coefficients (r) between the MRI Total Scale and Subscale Scale Scores^a

Scales	1	2	3	4
1. MRI O	–	.45**	.46**	.87**
2. MRI R		–	.34**	.82**
3. MRI C			–	.60**
4. MRI Total				–

Note. MRI O = Othering (Anti-Effeminacy and Homo-Negativity); MRI R = Responsibility (Dependability and Success); MRI C = Control (Dominance and Toughness). ^an = 319 after listwise deletion of missing data.

** p < .01

Table 5

Fisher's Transformation (z) of Correlation Coefficients between the MRI Total Scale and Subscale Scale Scores^a

MRI Scale Correlation	1	2	3	4	5	6
1. Othering-Responsibility (r = .45)	–	-.16	1.64	-10.66**	-8.45**	-2.62**
2. Othering-Control (r =.46)	–	–	1.80	-10.51**	-8.29**	-2.46*
3. Responsibility-Control (r =.34)	–	–	–	-12.31**	-10.09**	-4.26**
4. Othering-Total (r =.87)	–	–	–	–	2.22*	8.04**
5. Responsibility-Total (r =.82)	–	–	–	–	–	5.83**
6. Control-Total (r =.60)	–	–	–	–	–	–

Note. ^an = 319 after listwise deletion of missing data.

* p < .05

** p < .01

Table 6

Correlation Coefficients between the MRI and CRGS^a, PAQ-M^b Scale as well as MRNI Scale Scores^c

Scales	MRI			
	Othering	Responsible	Control	Total
GRCS Success/Power/Competition	.20*	.43**	.25*	.39**
Restrictive Emotionality	.27**	.02	.37**	.23*
Restrictive Affectionate Behavior btw Men	.49**	.08	.33**	.39**
Conflict btw Work and Family Relations	.12	.23*	.17	.22*
Total	.40**	.26**	.42**	.45**
PAQ Masculinity	-.10	.20*	.05	.06

MRNI	Avoidance of Femininity	.75**	.48**	.55**	.76**
	Rejection of Homosexuals	.75**	.45**	.49**	.73**
	Self-Reliance	.35**	.57**	.37**	.54**
	Aggression	.48**	.56**	.36**	.61**
	Achievement/Status	.62**	.56**	.47**	.71**
	Attitudes twd Sex	.61**	.57**	.44**	.70**
	Restrictive Emotionality	.57**	.49**	.67**	.68**
	Nontraditional Attitudes twd Masculinity	-.43**	-.44**	-.44**	-.54**
	Total	.77**	.66**	.63**	.87**

Note. ^a*n* = 102 after listwise deletion of missing data. ^b*n* = 105 after listwise deletion of missing data. ^c*n* = 104 after listwise deletion of missing data.

* *p* < .05

** *p* < .01