Intention to dropout and study satisfaction: testing item bias and structural invariance of measures for South African first-year university students

Karina Mostert, Clarisse van Rensburg and Reitumetse Machaba
Management Cybernetics Research Entity,
Faculty of Economic and Management Sciences, North-West University,
Potchefstroom, South Africa

Abstract
Purpose – This study examined the psychometric properties of intention to drop out and study satisfaction measures for first-year South African students. The factorial validity, item bias, measurement invariance and reliability were tested.
Design/methodology/approach – A cross-sectional design was used. For the study on intention to drop out, 1,820 first-year students participated, whilst 780 first-year students participated in the study on satisfaction with studies. Confirmatory factor analysis (CFA), differential item functioning (DIF), measurement invariance and internal consistency were used to test the scales.
Findings – A one-factor structure was confirmed for both scales. For the intention to drop out scale, Items 3 and 4 were identified with statistically significant item bias; however, these differences had no practical impact. Except for scalar invariance for language, sufficient measurement invariance was established. No problematic items were identified for the study satisfaction scale.
Practical implications – In essence, this study provides evidence of two short measures that are culturally sensitive that could be used as short and valid measures across contextual boundaries as practically valuable tools to measure intention to drop out and study satisfaction in diverse and multicultural contexts.
Originality/value – This study contributes to limited research on bias and invariance analyses for scales that can be used in interventions to identify students at risk of leaving the university and utilising psychometric analyses to ensure the applicability of these two scales in diverse and multicultural settings.

Keywords Intention to drop out, Study satisfaction, First-year university students, Factorial validity, Item bias, Measurement invariance, Internal consistency

Paper type Research paper

Introduction
The transitioning process of first-year university students is often regarded as a stressful experience (Van Zyl and Dhurup, 2018). This transition is particularly problematic in South
Africa and is associated with exceptionally high dropout rates (Van Zyl, 2017). In response to dealing with these challenges, two initiatives have been established in South Africa, the South African National Resource Centre (SANRC) for the First-Year Experience and Students in Transition and the Siyaphumelela (“We Succeed”) student success initiative. The SANRC aims to improve student success through the distribution and development of research regarding the first-year experience (Nyar, 2020), whilst the Siyaphumelela (“We Succeed”) initiative aims to expand evidence-based postsecondary student success strategies across Higher Education Institutions (HEIs) in South Africa. Based on the focus of these initiatives, two essential constructs that are imperative to consider in research on first-year students are the intention to drop out and study satisfaction.

Retention of students has been considered a quality indicator for many universities (Bernardo et al., 2022). Therefore, HEIs must identify students who intend to drop out and intervene before they leave university and do not return. The term intention to drop out can be described as a gradual process of goal disengagement (Ghassemi et al., 2017), where students’ conflict with a previous goal (i.e. to graduate from university) disengage with the goal and eventually abandon the goal (i.e. drop out of university) (Scheunemann et al., 2022). Study satisfaction refers to the extent to which students evaluate various aspects of their studies, such as their major, conditions of studies and having unfulfilled expectations (Scheunemann et al., 2022; Westermann and Heise, 2018) and can be conceptualised as the student’s level of satisfaction, general experience, or attitude towards their academic studies or the university (Duque, 2014).

Several studies investigated the relationship between the intention to drop out and study satisfaction. Scheunemann et al. (2022) position intention to dropout as a mediator between internal and external causes of student dropout and actual dropout. They viewed study satisfaction as a possible determinant of the intention to drop out. Their three-wave longitudinal study results showed a dynamic interplay between variables in the dropout process and showed that high dropout intention is significantly related to study satisfaction. Similarly, Bernardo et al. (2022) position study satisfaction and expectations with the course of study as predictors of intention to drop out. Their findings emphasise that multiple variables influence intentions to drop out directly and indirectly. These findings align with Duque’s (2014) conceptual framework from a literature synthesis on the relationship between students’ satisfaction, perceived learning outcomes and dropout intentions.

Our study positions intention to drop out and study satisfaction slightly differently than the studies mentioned above since the study forms part of a larger research project called StudyWell: Student Well-being and Success. The StudyWell project utilises the leading approach in occupational health and well-being research, the Job Demands-Resources (JD-R) theory (c.f. Bakker and Demerouti, 2017; Bakker et al., 2023). One of the assumptions of JD-R theory is that two processes underly well-being. The health-impairment process occurs when individuals experience severe demands, which may lead to exhaustion, health problems and unfavourable outcomes for the organisation, such as employee turnover. The motivational process occurs when resources are available to deal with the effect of demands and foster creativity and motivation, such as employee engagement, which may lead to positive outcomes for the individual and organisation (e.g. good performance). This approach enables the investigation of theoretically and empirically neglected reciprocal relations with the negative and positive outcomes of students’ health-impairment and motivational processes (such as intention to drop out and study satisfaction). As such, integrating the streams of dropout literature with an integrated well-being theory, such as JD-R theory, may allow linking different aspects of students’ lives (their demands, resources and well-being) to essential outcomes for the student and university (Duque, 2014; Scheunemann et al., 2022).
Periodic assessments are needed to accurately establish and measure students’ intention to drop out of the university and their satisfaction with their studies (Bacila et al., 2014). However, psychological testing is governed in South Africa by the Employment Equity Act No. 55 of 1998, Section 8 (President of the Republic of South Africa, 1998). According to this Act, assessments are prohibited unless they can scientifically be proven reliable and valid, can be applied fairly to all ethnic groups and cultures and are not biased against any person or group.

Item bias refers to the event in which the meaning of an item, or multiple items, is understood identically across different cultures or groups and relates to item-level irregularities. An item is biased when score differences do not occur based on actual differences in the measured underlying construct but because of item-level incongruities (Van de Vijver and Tanzer, 2004).

Establishing the configural invariance of measures is essential to investigate if the factor structure fits the data equally in all groups (i.e. has the same pattern across sub-groups). Configural invariance shows to what extent the factor structure can be replicated in the same way across different groups. Metric invariance is an essential property of a scale that indicates whether each unit of measurement (i.e. each item) contributes equally to the latent construct across different sub-groups. Scalar invariance refers to establishing whether a test score has the same meaning in terms of how it is interpreted, regardless of the cultural background (Van de Vijver and Tanzer, 2004).

In addition to adhering to legislation, establishing measurement invariance is also essential for practical reasons because inaccurate assessment may influence the valid interpretation and correct estimation of effects in research (Teresi and Fleishman, 2007). Many decisions are made on individual and group differences. Ensuring equivalent measurement is essential before making comparisons because a lack of measures’ equivalence (or invariance) makes group comparisons ambiguous (Gregorich, 2006; Teresi and Fleishman, 2007). As a result, flawed instruments may lead to suboptimal decisions (Teresi and Fleishman, 2007) and may impact policy planning and implementation of interventions (Perkins et al., 2006).

This study aimed to test the psychometric properties of two short measures, intention to drop out and study satisfaction, to establish whether these measures are valid, reliable, unbiased and invariant for different language, campus and gender groups in a sample of first-year university students in South Africa.

Method
Research procedure and participants
An ethical application was submitted and approved and a formal ethics number was obtained. The goal and purpose, confidentiality and anonymity regarding personal information and the possible value to the university and students were explained. Emphasis was placed on participation being voluntary. The data collection was part of the larger StudyWell project, where intention to drop out was included in one study (Study 1) as an outcome of the health-impairment process and study satisfaction was included in another study (Study 2) as an outcome of the motivational process as described in JD-R theory (Bakker and Demerouti, 2017; Bakker et al., 2023).

Data were collected from the three campuses of the university. The university was formed by merging a historically black university and a historically white university as part of the South African government’s plan to transform higher education. The merger formed three campuses, each with a unique and diverse culture hosting students from different cultures and language groups.

The sample in Study 1 consisted of 1820 research participants between the ages of 17 and 24. In terms of language, 39% were Afrikaans, followed by Setswana (27%), Sesotho (9.2%)
and English (7.3%). The remaining 14.8% of the sample consisted of participants who spoke one of the eleven official languages in South Africa or another language. The most significant number of participants (53.8%) studied at campus 2, followed by 28.2% of students who studied at campus 1 and 17.3% studying at campus 3. Most research participants were female (65.2%; males were 33.7%). The sample in Study 2 consisted of 780 research participants, of whom the majority were between 18 and 20 years old (73.7%). Regarding language, 38.8% indicated that they spoke Afrikaans, 33.1% indicated that they spoke Setswana and 6.2% indicated that they spoke Sesotho, three of the 11 official languages in South Africa. Most of the total sample was studying at either campus 2 (50.5%) or campus 1 (38.3%), with the smallest number of participants studying at campus 3 (9.7%). Concerning gender, the sample comprised 61.8% female and 38.2% male participants.

Measuring instruments

Intention to drop out

The work-related scale of intention to leave the organisation, developed by Sjöberg and Sverke (2000), was adapted to measure intention to drop out for the student context (“If it was up to me, I would quit my studies and do what I want”; “I feel that I want to leave the university before I finish my studies”; “I want to quit my studies”; and “If I was completely free to choose I would leave the university and find a job”). All items are scored on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Sjöberg and Sverke (2000) confirmed an internal consistency of the scale, obtaining a Cronbach’s alpha coefficient of 0.83.

Study satisfaction

The job satisfaction scale, developed by Hellgren et al. (1997), was adapted to measure study satisfaction. The work-related scale originally consisted of three items and a fourth item was added. These four items were adapted to fit the student context (i.e. “I enjoy my studies”; “I am content with my studies”; “I am satisfied with my studies”; and “I am happy in my studies”). The scale was scored on a five-point Likert-type scale that ranges from 1 (Strongly disagree) to 5 (Strongly agree). Hellgren et al. (1997) confirmed the scale’s internal consistency, obtaining a Cronbach’s alpha coefficient of 0.86.

Statistical analysis

Mplus 8.6 (Muthén and Muthén, 2021) was used to conduct the statistical analyses. Confirmatory factor analysis (CFA) was used to test factorial validity and invariance. Maximum likelihood estimation was used, with the covariance matrix as input (Muthén and Muthén, 2014). The following fit indices were considered to assess the models’ goodness-of-fit: the $\chi^2$ statistic, the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA) and the standardised root mean square residual (SRMR). For CFI and TLI, the acceptable fit is considered at 0.90 and above (Byrne, 2001; Hoyle, 1995). A cut-off value of 0.05 or less indicates a good fit for RMSEA, whereas values between 0.05 and 0.08 are considered a good model fit (Chen et al., 2008). The guidelines of DiStefano et al. (2009) were followed to interpret the factor loadings of items.

Differential item functioning (DIF) was used to test for item bias. The lordif package (Choi et al., 2011) in RStudio (https://www.rstudio.com/) was used. The following formulae were used and compared to test for uniform and non-uniform bias, using ordinal logistic regression to generate three likelihood-ratio $\chi^2$ statistics (Choi et al., 2011):

Model 0: $\text{logit} \ P(u_i \geq k) = \alpha_k$

Model 1: $\text{logit} \ P(u_i \geq k) = \alpha_k + \beta_1 * \text{ability}$
Model 2: logit \( P(u_i \geq k) = \alpha_k + \beta_1 \cdot \text{ability} + \beta_2 \cdot \text{group} \)

Model 3: logit \( P(u_i \geq k) = \alpha_k + \beta_1 \cdot \text{ability} + \beta_2 \cdot \text{group} + \beta_2 \cdot \text{ability} \cdot \text{group} \)

Bias can be identified when statistically significant differences are detected by comparing the log-likelihood values of models (\( p < 0.01 \)). Uniform bias can be detected when comparing logistic Models 1 and 2 (\( \chi^2_{12} ; \text{df} = 1 \)), whereas non-uniform bias can be detected when comparing logistic Models 2 and 3 (\( \chi^2_{23} ; \text{df} = 2 \) (Choi et al., 2011). The pseudo-McFadden \( R^2 \) statistic was used to test the magnitude of the DIF. The magnitude of DIF can be classified as negligible (<0.13), moderate (between 0.13 and 0.26) and large (>0.26) (Zumbo, 1999).

Additionally, using the difference in the \( \beta_1 \) coefficient from Models 1 and 2, the practical significance of uniform DIF was determined with 10% differences between Models 1 and 2, indicating a practically meaningful effect (Crane et al., 2004). Lower thresholds of 5% and even 1% are also used (Crane et al., 2007). A threshold of 5% was used in this study.

Measurement invariance was investigated based on language (Afrikaans, Sesotho, Setswana and English), campus (three campuses) and gender (male and female). Multigroup analysis was used that includes the: (1) configural invariance model (i.e. the baseline model for the more constrained models and the test if the factor structure is analogous across groups) (2) metric invariance model (assumes similarity or invariance of the factor loading across different groups); and (3) scalar invariance model (tests if the factor loadings and item intercepts similar or invariant across different groups) (Preti et al., 2013). CFI and RMSEA were used as cut-off points. The CFI value is considered adequate if values are above 0.90 and better if they are higher than 0.95. With regards to RMSEA, the cut-off value is < 0.08; better is < 0.05 (Van De Schoot et al., 2012); however, as recommended by Shi et al. (2019), changes in CFI (\( \Delta \text{CFI} \)) were used. Therefore, a \( \Delta \text{CFI} \) value higher than 0.01 between two nested models indicates that the added group constraints have led to a poorer fit, i.e. the more constrained model was rejected. The loadings of items were freed to achieve partial metric invariance (Preti et al., 2013).

Cronbach’s alpha coefficient was used to determine the reliability of the scales. Cronbach’s alpha reliability coefficient typically ranges between 0 and 1. George and Mallery (2003) provide the following rules of thumb to interpret the values of Cronbach’s alpha coefficients: \( \alpha > 0.90 \) – Excellent, \( \alpha > 0.80 \) – Good, \( \alpha > 0.70 \) – Acceptable, \( \alpha > 0.60 \) – Questionable, \( \alpha > 0.50 \) – Poor and \( \alpha < 0.50 \) – Unacceptable.

Results

Psychometric analyses for the intention to drop out scale

Factorial validity. The results showed that a one-factor model was an excellent fit to the data (\( \chi^2 = 8.723, \text{df} = 2.000, \text{CFI} = 0.988, \text{TLI} = 0.965, \text{RMSEA} = 0.058, \text{SRMR} = 0.019 \)). The factor loadings are presented in Table 1 below. All items had acceptable and statistically significant factor loadings (\( \lambda \)) ranging from 0.66 to 0.88.

<table>
<thead>
<tr>
<th>Item</th>
<th>Item text</th>
<th>Loading</th>
<th>S.E.</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>If it was up to me, I would quit my studies and do what I want</td>
<td>0.669</td>
<td>0.028</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 2</td>
<td>I feel that I want to leave the university before I finish my studies</td>
<td>0.886</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 3</td>
<td>I want to quit my studies</td>
<td>0.867</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 4</td>
<td>If I was completely free to choose I would leave the university and find a job</td>
<td>0.757</td>
<td>0.027</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note(s): S.E. = standard error and all \( p \)-values <0.001
Source(s): Authors’ own work

Table 1. Standardised factor loadings for the latent variables of the intention to drop out scale
**Item bias.** Uniform, non-uniform and total bias were tested for the intention to drop out scale. The results are presented in Table 2.

First, items are flagged when statistically significant differences are detected (items marked in italic in Table 2). As shown in Table 2, Item 3 displayed language and campus-related DIF, whilst Item 4 displayed DIF for language, campus and gender groups. Four visual graphs per item are provided below that display additional diagnostic information to interpret the bias detected in these items. The upper-left graph shows the item characteristic curves for the different sub-groups (i.e. different language, campus or gender groups). The lower-left graph shows the item response functions for the sub-group parameter estimates (slope and category threshold values for each sub-group). The upper-right graphs display the absolute difference between the item characteristic curves of the different groups. The lower-right graph shows the absolute difference between the item characteristic curves of the sub-groups weighted by the score distribution (Choi et al., 2011).

Table 2 and Figure 1 show that Item 3 displayed uniform, non-uniform and total bias for the different language groups. The top left plot in Figure 1 shows that the slope of the function for the Afrikaans groups was slightly higher than for the other language groups. It can also be seen in the bottom left plot that the category threshold values for the Afrikaans groups are noticeably different compared to the other groups. The top right plot shows a difference in the item-true-score functions; however, this difference is negligible, as seen in the density-weighted impact (bottom right plot). Based on this information, pseudo-McFadden $R^2$ statistic values < 0.13 and the difference in the $\beta_1$ coefficient smaller than 5%, DIF’s magnitude or practical impact for Item 3 can be classified as negligible. Similarly, Item 4 displayed uniform and total bias in Figure 2. Although noticeable differences can be seen between groups in the graphs, these differences also have no practical impact with pseudo-McFadden $R^2$ statistic values < 0.13 and $\Delta \beta_1$ coefficient smaller than 5%.

Regarding campus, Items 3 and 4 were flagged as items with statistically significant biased items; Item 3 with uniform and total bias and Item 4 with uniform, non-uniform and total bias. Some differences between the campus groups can be seen in the plots (in Figures 3 and 4), specifically Campus 1 (dark black line) scoring somewhat higher or lower than the other groups. However, regarding the magnitude of these items, the density-weighted impact

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>$\chi^2_{12}$</th>
<th>$\chi^2_{13}$</th>
<th>$\chi^2_{23}$</th>
<th>$\Delta \beta_1$</th>
<th>$R^2_{12}$</th>
<th>$R^2_{13}$</th>
<th>$R^2_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Item 1</td>
<td>0.0659</td>
<td>0.7469</td>
<td>0.2088</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0022</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.9110</td>
<td>0.2701</td>
<td>0.6151</td>
<td>0.0012</td>
<td>0.0002</td>
<td>0.0016</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0200</td>
<td>0.0071</td>
<td>0.0142</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.0000</td>
<td>0.1753</td>
<td>0.0000</td>
<td>0.0185</td>
<td>0.0096</td>
<td>0.0111</td>
<td>0.0016</td>
</tr>
<tr>
<td>Campus</td>
<td>Item 1</td>
<td>0.8373</td>
<td>0.8219</td>
<td>0.9453</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.7608</td>
<td>0.7933</td>
<td>0.9083</td>
<td>0.0010</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.0001</td>
<td>0.9261</td>
<td>0.0007</td>
<td>0.0201</td>
<td>0.0074</td>
<td>0.0075</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.0000</td>
<td>0.0072</td>
<td>0.0000</td>
<td>0.0197</td>
<td>0.0104</td>
<td>0.0133</td>
<td>0.0029</td>
</tr>
<tr>
<td>Gender</td>
<td>Item 1</td>
<td>0.4716</td>
<td>0.1180</td>
<td>0.2274</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.0008</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.8245</td>
<td>0.3366</td>
<td>0.6149</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0003</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.9482</td>
<td>0.8942</td>
<td>0.9891</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.0011</td>
<td>0.2040</td>
<td>0.0022</td>
<td>0.0046</td>
<td>0.0031</td>
<td>0.0036</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Table 2.** Summary of the DIF analyses for the intention to drop out scale

**Note(s):** $\chi^2_{12}$ = chi-square of model 1 compared to model 2; $\chi^2_{13}$ = chi-square of model 1 compared to model 3; $\chi^2_{23}$ = chi-square of model 2 compared to model 3; $\Delta \beta_1$ = change in beta coefficient; $R^2_{12}$ = pseudo-McFadden $R^2$ of model 1 compared to model 2; $R^2_{13}$ = pseudo-McFadden $R^2$ of model 1 compared to model 3 and $R^2_{23}$ = pseudo-McFadden $R^2$ of model 2 compared to model 3

**Source(s):** Authors’ own work
Figure 1. Graphical display of Item 3 with respect to language
Source(s): Authors’ own work

Figure 2. Graphical display of Item 4 with respect to language
Source(s): Authors’ own work

Dropout and study satisfaction
Figure 3. Graphical display of Item 3 with respect to campuses

Source(s): Authors’ own work

Figure 4. Graphical display of Item 4 with respect to campuses

Source(s): Authors’ own work
seen in the bottom right plot, as well as pseudo-McFadden $R^2$ statistic values < 0.13 and $\Delta \beta_1$ coefficient smaller than 5%, indicate that the practical significant effect is, again, negligible.

For male and female students, Item 4 showed statistically significant bias. The item-true-score functions (upper-left graph) show that male students are prone to endorse Item 4 with higher categories compared to female students with the same overall intention to drop out. Again, as can be seen by the weighted by density impact, this effect is barely noticeable and, therefore, negligible (see Figure 5).

**Measurement invariance.** Table 3 shows the measurement invariance across the language, campus and gender groups included for the intention to drop out scale.

Table 3 shows that the intention to drop out scale was invariant regarding configural, metric and scalar invariance for language, campus and gender groups, except scalar

Source(s): Authors’ own work

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>$\Delta$CFI</th>
<th>RMSEA</th>
<th>$\Delta$RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Configural</td>
<td>24.39</td>
<td>8</td>
<td>0.984</td>
<td>–</td>
<td>0.074</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>37.30</td>
<td>17</td>
<td>0.980</td>
<td>–0.004</td>
<td>0.056</td>
<td>–0.018</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>73.13</td>
<td>26</td>
<td>0.953</td>
<td>–0.027</td>
<td>0.069</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Partial scalar</td>
<td>62.94</td>
<td>25</td>
<td>0.974</td>
<td>–0.006</td>
<td>0.057</td>
<td>0.001</td>
</tr>
<tr>
<td>Campus</td>
<td>Configural</td>
<td>99.69</td>
<td>54</td>
<td>0.993</td>
<td>–</td>
<td>0.072</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>121.03</td>
<td>62</td>
<td>0.991</td>
<td>–0.002</td>
<td>0.076</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>147.30</td>
<td>79</td>
<td>0.989</td>
<td>–0.002</td>
<td>0.073</td>
<td>–0.003</td>
</tr>
<tr>
<td>Gender</td>
<td>Configural</td>
<td>30.47</td>
<td>4</td>
<td>0.984</td>
<td>–</td>
<td>0.074</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>39.54</td>
<td>7</td>
<td>0.970</td>
<td>–0.006</td>
<td>0.072</td>
<td>–0.014</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>48.84</td>
<td>10</td>
<td>0.965</td>
<td>–0.005</td>
<td>0.066</td>
<td>–0.006</td>
</tr>
</tbody>
</table>

Note(s): $\chi^2$ = chi-square; df = degrees of freedom; CFI = comparative fit index; $\Delta$CFI = delta (change in) CFI; RMSEA = Root mean square error of approximation; $\Delta$RMSEA = delta (change in) RMSEA

Source(s): Authors’ own work

Figure 5. Graphical display of Item 4 with respect to gender

**Table 3.** Summary of measurement invariance analyses for the intention to drop out scale
invariance for language groups. The $\Delta$CFI value $> 0.01$ between the two nested models showed that scalar invariance could not be confirmed for language groups. Partial scalar invariance was achieved by releasing the intercepts for Items 3 and 4 in the Afrikaans group and the intercept of Item 4 in the other groups.

Internal consistency. A Cronbach’s alpha coefficient of 0.85 demonstrated acceptable reliability ($\alpha \geq 0.70$) for the intention to drop out scale.

Psychometric properties of the study satisfaction scale

Factorial validity. The results showed a good fit to the data ($\chi^2 = 0.646; \text{df} = 2; \text{CFI} = 1.000; \text{TLI} = 1.000; \text{RMSEA} = 0.000; \text{SRMR} = 0.004$). Table 4 shows the results for the standardised loadings of the items for the latent variables of the scale. All items had acceptable and statistically significant factor loadings ($\lambda$) ranging from 0.753 to 0.870.

Item bias. DIF analyses were used to test for item bias. The results are shown in Table 5. No uniform or non-uniform bias was found in the items of the study satisfaction scale across the different language, campus and gender groups. In addition, the changes in the beta coefficients across all groups were well below the 5% cut-off set for this study, demonstrating that the items are not biased across the different groups.

Measurement invariance. Measurement invariance (configural, metric and scalar) was tested between the different language, campus and gender groups. Table 6 shows the results.

### Table 4.
Standardised factor loadings for the latent variables of the study satisfaction scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Item text</th>
<th>Loading</th>
<th>S.E.</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>I enjoy my studies</td>
<td>0.842</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 2</td>
<td>I am content with my studies</td>
<td>0.753</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 3</td>
<td>I am satisfied with my studies</td>
<td>0.858</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Item 4</td>
<td>I am happy in my studies</td>
<td>0.870</td>
<td>0.022</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Note(s):** S.E. = standard error and all $p$-values $< 0.001$

**Source(s):** Authors’ own work

### Table 5.
Summary of the DIF analyses for the study satisfaction scale

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>$\chi^2_{12}$</th>
<th>$\chi^2_{13}$</th>
<th>$\chi^2_{23}$</th>
<th>$\Delta \beta_1$</th>
<th>$R^2_{12}$</th>
<th>$R^2_{13}$</th>
<th>$R^2_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Item 1</td>
<td>0.1657</td>
<td>0.2148</td>
<td>0.3327</td>
<td>0.0076</td>
<td>0.0027</td>
<td>0.0044</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.1899</td>
<td>0.2798</td>
<td>0.4165</td>
<td>0.0055</td>
<td>0.0025</td>
<td>0.0039</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.1691</td>
<td>0.2707</td>
<td>0.4468</td>
<td>0.0078</td>
<td>0.0027</td>
<td>0.0039</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.1811</td>
<td>0.3870</td>
<td>0.6950</td>
<td>0.0131</td>
<td>0.0026</td>
<td>0.0032</td>
<td>0.0006</td>
</tr>
<tr>
<td>Campus</td>
<td>Item 1</td>
<td>0.8639</td>
<td>0.7468</td>
<td>0.4388</td>
<td>0.0016</td>
<td>0.0002</td>
<td>0.0013</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.9440</td>
<td>0.4402</td>
<td>0.1620</td>
<td>0.0011</td>
<td>0.0001</td>
<td>0.0026</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.8622</td>
<td>0.9301</td>
<td>0.7540</td>
<td>0.0013</td>
<td>0.0002</td>
<td>0.0006</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.8573</td>
<td>0.9717</td>
<td>0.9003</td>
<td>0.0014</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>Item 1</td>
<td>0.4462</td>
<td>0.7468</td>
<td>0.9524</td>
<td>0.0022</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 2</td>
<td>0.1131</td>
<td>0.2792</td>
<td>0.8405</td>
<td>0.0004</td>
<td>0.0014</td>
<td>0.0015</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 3</td>
<td>0.6203</td>
<td>0.8559</td>
<td>0.7978</td>
<td>0.0016</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>Item 4</td>
<td>0.1608</td>
<td>0.2733</td>
<td>0.4282</td>
<td>0.0045</td>
<td>0.0012</td>
<td>0.0016</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

**Note(s):** $\chi^2_{12}$ = chi-square of model 1 compared to model 2; $\chi^2_{13}$ = chi-square of model 1 compared to model 3; $\chi^2_{23}$ = chi-square of model 2 compared to model 3; $\Delta \beta_1$ = change in beta coefficient; $R^2_{12}$ = pseudo-McFadden $R^2$ of model 1 compared to model 2; $R^2_{13}$ = pseudo-McFadden $R^2$ of model 1 compared to model 3 and $R^2_{23}$ = pseudo-McFadden $R^2$ of model 2 compared to model 3.

**Source(s):** Authors’ own work
The results in Table 6 show that the study satisfaction scale has configural, metric and scalar invariance across the different language, campus and gender groups, with CFI scores ranging from 0.985 to 1.000. This indicates strong measurement invariance (Van De Schoot et al., 2012).

**Internal consistency.** Cronbach’s coefficient alpha was 0.90, indicating acceptable reliability ($\alpha \geq 0.70$).

**Discussion and practical implications**

The results showed that a one-factor model for each scale represented an excellent fit to the data. Regarding item bias, Items 3 and 4 of the intention to drop out scale showed some statistically significant bias. However, these differences were negligible and had no practical impact or effect. No bias was detected in any of the study satisfaction scales’ items. Configural, metric and scalar invariance were tested. Although the intercepts for Items 3 and 4 in one language group had to be released to reach scalar invariance for the intention to drop out scale (implying that means can still be compared based on language group, if required), the findings indicate that both scales have configural invariance (same one-factor structure), metric invariance (similar factor loadings) and scalar invariance (similar intercepts) across the different groups. Both scales also demonstrated good internal consistency.

These results emphasise the importance for HEIs to invest in the multicultural assessment of measures in cross-cultural settings. Even though South Africa is a very diverse country where multicultural assessment is guided by legislation, migration and globalisation are a reality for many HEIs worldwide (Maringe and Foskett, 2010). There has been a significant upsurge in the number of international students to HEIs in many countries (IIE Open Doors / Enrollment Trends, 2020) and has created linguistically and culturally diverse student groups that give rise to various opportunities for cultural constructions and re-constructions (Wang and Sun, 2022). Using measures that take cultural factors into account, could contribute to credible practices that are rigorous, unbiased, have interpretive power and enable accurate interpretation and intervention for student success initiatives (Lacko et al., 2022).

Few measures have been established to assist with dropout preventative interventions (Bernardo et al., 2022), specifically for diverse settings. The two scales presented in this study could be used as short and valid measures across contextual boundaries and can be used as practically valuable tools to measure intention to drop out and study satisfaction in diverse and multicultural contexts. In addition, investment in student success initiatives and interventions at tertiary levels should ideally be transferred to students’ employability.

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>$\Delta$CFI</th>
<th>RMSEA</th>
<th>$\Delta$RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Configural</td>
<td>7.91</td>
<td>8</td>
<td>1.000</td>
<td></td>
<td>0.000</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>21.30</td>
<td>17</td>
<td>0.994</td>
<td>-0.006</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>34.90</td>
<td>26</td>
<td>0.988</td>
<td>-0.006</td>
<td>0.046</td>
<td>0.006</td>
</tr>
<tr>
<td>Campus</td>
<td>Configural</td>
<td>18.55</td>
<td>6</td>
<td>0.985</td>
<td></td>
<td>0.091</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>16.80</td>
<td>12</td>
<td>0.994</td>
<td>0.004</td>
<td>0.040</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>22.48</td>
<td>18</td>
<td>0.994</td>
<td>0.000</td>
<td>0.031</td>
<td>0.009</td>
</tr>
<tr>
<td>Gender</td>
<td>Configural</td>
<td>6.75</td>
<td>4</td>
<td>0.996</td>
<td></td>
<td>0.042</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Metric</td>
<td>12.93</td>
<td>7</td>
<td>0.991</td>
<td>-0.005</td>
<td>0.047</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Scalar</td>
<td>17.75</td>
<td>10</td>
<td>0.989</td>
<td>-0.002</td>
<td>0.045</td>
<td>-0.002</td>
</tr>
</tbody>
</table>

**Note(s):** $\chi^2$ = chi-square; df = degrees of freedom; CFI = comparative fit index; $\Delta$CFI = delta (change in) CFI; RMSEA = Root mean square error of approximation and $\Delta$RMSEA = delta (change in) RMSEA

**Source(s):** Authors’ own work

Table 6. Summary of measurement invariance analyses for the study satisfaction scale
employment and general functioning after graduation. From an institutional perspective, it is essential to track graduates’ employment destinations and functioning as graduates in a continual cycle, from the time students enter university until they exit, to fine-tune and improve intervention effectiveness where necessary (Jackson et al., 2013). To accomplish this, a fine-grained and aligned implementation of a questionnaire methodology is necessary (Manathunga et al., 2009). Since intention to drop out and study satisfaction are variables similar to the work-related concepts of intention to leave the organisation and job satisfaction, two widely used scales in occupational psychology have been adapted for the student context in our study. The advantage of this approach is to have systematic stability between the questionnaires administered for students vs graduates.

Limitations and recommendations
Our study had several limitations that should be mentioned and provides ideas for future research. Because this study was part of a larger initiative for first-year students, the results apply specifically to South African students. Another limitation concerns the language groups included in our sample. This limits international generalisations and generalisations to other language groups in South Africa, which has 11 official languages (Statistics South Africa, 2018). Future researchers should include samples representing other language groups in South Africa or English as a language group for cross-cultural comparisons.

Although using scales developed to measure intention to leave and job satisfaction (work-related scales) and adapting them for the student context can be beneficial (as explained above), the questions can seem too straightforward and present-generation students may not express their true feelings on such questions. Future research could explore redesigning the questionnaires by asking the questions more indirectly to obtain true intentions and feelings. These scales were chosen because they are short and concise, characteristics that are beneficial when students have to complete long questionnaires. However, future research could include scales explicitly designed for students that are more comprehensive and could enable researchers to link student motivations as a precursor to their ultimate actions (e.g. as outlined in the studies of Bernardo et al., 2022, Duque, 2014 and Scheunemann et al., 2022).

Finally, we used a cross-sectional design and two different samples. As a result, the relationship between turnover intention and study satisfaction could not be examined. Future researchers should explore how the intention to drop out and study satisfaction scales fit within the larger nomological net of first-year university students.

References


**Further reading**


About the authors
Karina Mostert is Professor in Industrial Psychology in the Faculty of Economic and Management Sciences at the North-West University, Potchefstroom Campus, South Africa. She conducted research on the topic of Occupational Health and Well-Being and focussed on subjective well-being, burnout, engagement, work–home interference and strengths use. Her research focus shifted to the health and well-being of university students, with a specific focus on their experiences, engagement and psychological resources that can assist in optimal functioning and performance. She is leading the project “StudyWell: Student Well-Being and Success” at the North-West University, South Africa. The project aims to develop a valid, reliable, culturally sensitive online data informed monitoring tool for student well-being, informed by in-depth qualitative investigation, to assess and proactively monitor the study climate, individual traits, states and behaviour of students to inform targeted and cost-effective interventions. Karina Mostert is the corresponding author can be contacted at: Karina.Mostert@nwu.ac.za

Clarisse van Rensburg is currently completing her Industrial Psychology internship at a retail company in South Africa and had previously worked in higher education at Varsity College as a senior planning and scheduling coordinator. She has completed her Master’s degree in Industrial and Organisational Psychology at the North-West University (Potchefstroom Campus), South Africa. Her research interests are in life satisfaction, first-year students and psychometrics.

Reitumetse Machaba is currently Organisational Development Consultant at Sanlam Corporate and had previously worked in higher education at North-West university focussing on organisational development and organisational culture initiatives. She has completed her Master’s degree in Industrial Psychology at the North-West University (Potchefstroom Campus), South Africa. Her research interests are in first-year students and psychometrics.