1          The title

2     Erin M. Buchanan[1], Jeffrey Pavlacic[2], Becca Huber[3], & Alyssa Counsell[4]

3          [1] Harrisburg University of Science and Technology

4          [2] University of Mississippi

5          [3] Idaho State University

6          [4] Ryerson University

7          Author Note

8     Add complete departmental affiliations for each author here. Each new line herein

9     must be indented, like this line.

10     Enter author note here.

11     Correspondence concerning this article should be addressed to Erin M. Buchanan, 326

12     Market St., Harrisburg, PA 17101. E-mail: ebuchanan@harrisburgu.edu

13                                            Abstract

14   One or two sentences providing a **basic introduction** to the field, comprehensible to a

15   scientist in any discipline.

16        Two to three sentences of **more detailed background**, comprehensible to scientists

17   in related disciplines.

18        One sentence clearly stating the **general problem** being addressed by this particular

19   study.

20        One sentence summarizing the main result (with the words "**here we show**" or their

21   equivalent).

22        Two or three sentences explaining what the **main result** reveals in direct comparison

23   to what was thought to be the case previously, or how the main result adds to previous

24   knowledge.

25        One or two sentences to put the results into a more **general context**.

26        Two or three sentences to provide a **broader perspective**, readily comprehensible to

27   a scientist in any discipline.

28        *Keywords:* keywords

29   Word count: X

<sup>30</sup>                                               The title

<sup>31</sup>     MAKE A PACKRAT - figure out how to reference other markdowns

<sup>32</sup>     Classify by sample size, classify by scale size random versus not random (drop paper)
<sup>33</sup> See if/where it breaks - configural - metric - scalar - strict residuals

<sup>34</sup>     (Buchanan, 2002) says that stuff is stuff and stuff. (**???**) said more stuff. This file
<sup>35</sup> needs an update.

<sup>36</sup>     ##Scale Development

<sup>37</sup>     Scale development typically begins with an underlying latent construct of interest to
<sup>38</sup> the researcher (DeVellis citation). (Olson 2010 citation) suggests compiling a list of items
<sup>39</sup> thought to represent the construct of interest. These items are generally evaluated by
<sup>40</sup> experts in the specified area. This concept is termed as construct or face validity (citation on
<sup>41</sup> construct or face validity). After eliminating problematic questions, data are collected and
<sup>42</sup> examined through classical test methods, specifically exploratory and confirmatory factor
<sup>43</sup> analysis (EFA; CFA; Worthington citation) or Item Response Theory methods (cite any IRT
<sup>44</sup> article or probably a seminal IRT article). Generally, EFA and CFA examine correlations
<sup>45</sup> between items in regards to how they relate to a latent construct, also termed factor.
<sup>46</sup> Specific items will sometimes correlate to form multiple factors, or psychological constructs,
<sup>47</sup> depending on the measure of interest. For example, the DASS (cite DASS or literally any
<sup>48</sup> other scale with multiple factors) will lead to clusters of items correlated with different
<sup>49</sup> underlying constructs (depression, anxiety, stress). Unfortunately, researchers will sometimes
<sup>50</sup> make questionable decisions in utilizing factor-analytic methods that may lead to
<sup>51</sup> misinterpreted measures and drastic clinical implications (mainly, an inability of a measure
<sup>52</sup> to actually measure what it is supposed to measure, which could lead to misdiagnosed or
<sup>53</sup> incorrectly diagnosed patients. I'm sure that there are other clinical implications too, but I

think that meaning in life measures are generally used as a supplement to therapy. If we use measures as a supplement to therapy, like meaning in life measures, then they could inhibit our ability to enhance therapy). Specifically, Preacher & MacMallum (2002 I think – check year) showcase common mistakes made in conducting factor analyses and how they could be avoided. Specifically, researchers will mistakenly use Principal Components Analysis (PCA) and will keep eigenvalues that are greater than 1. Additionally, they use varimax rotation (Kaiser, 1970: explain why using these things is problematic to research. Need to look into this more because I honestly don't know much about the different types of rotations). Initially, EFA examines factor loadings in order to examine how well items correlate with a latent construct(s). When adequate factor loadings (> .300) have been achieved for each item, CFA assesses model fit by means of fit statistics (cite whoever came up with the .300 criteria). Scales demonstrating adequate model fit are then published for use in psychological assessment. Obviously, scale development is crucial and of utmost importance in accurately measuring constructs of interest.

**Multi-Group Confirmatory Factor Analysis**

A topic of interest to many researchers is whether or not measures yield different results when conducting factor analyses across groups. Generally, it is necessary to examine whether or not the psychometrically-supported factor structure holds across different groups of interest. This is done by examining measurement invariance, which tells us whether or not our measure will yield the same attributes across such groups (Beaujean, 2014). Multi-group Confirmatory Factor Analysis (MGCFA) applies CFA principles to different groups (usually two – more than two would warrant a different type of analyses such as MIMIC) across a measure or measures of interest. As an example of a type of multi-group analysis, this study sought to examine scale-delivery type (random vs. not random) in order to conclude whether or not randomization of different meaning in life measures produces different results in

79 regards to factor structure of said scales. Specifically, we initially utilized a stepwise

80 approach to examine model fit across all groups (both random and non-random). Then, data

81 were into both random and non-random data in order to examine model fit across these

82 individual groups. Each group provides the researcher with a set of fit indices by which to

83 examine model fit. The rationale for conducting such an analysis is to discover whether or

84 not each group produces less than desired fit statistics. With this information, the researcher

85 is able to tell whether or not different group reports differently on the given scale.

86 Regardless of model fit, we continued with the suggested stepwise approach by calculating

87 different types of invariances. Each level of invariance adds certain restrictions to the model.

88 Configural invariance, also referred to as "equal form invariance," allows the researcher

89 to understand whether or not factor structure/loadings are identical across specified groups.

90 Regardless of whether or not equal forms was supported across groups, it is then suggested to

91 analyze data using metric invariance. Metric invariance, too, examines factor loadings across

92 groups. This analysis is supported if this test of invariance did not differ significantly from

93 configural invariance. (CFI difference must be less $< .01$). Assuming our model does not

94 significantly differ in CFI after examining metric invariance, researchers generally conduct an

95 analysis of scalar invariance. Scalar invariance examines indicator intercepts and determines

96 whether or not these are equal across groups. Additionally, scale invariance determines

97 whether or not group membership influences raw scores. As with metric invariance, if the

98 change in CFI is less than .01, this assumption has been met. Finally, assuming that the

99 metric invariance assumption has been met, researchers will then examine partial invariance.

100 If the model breaks down, researchers will generally debate on whether or not to abandon

101 the measure or remove invariant items for further analyses, as noninvariant items may affect

102 construct validity as well as theory behind the measure of interest (Cheung & Rensvold,

103 1999). Our analyses present noninvariant questions across random/non-random groups so

104 that researchers are better able to decide whether or not it would be prudent to randomize

105 question order. This section presents a summary of MGCFA and the rationale for its usage.

106  It is important to conduct this type of analysis in order to improve construct validity.

**Meaning in Life**

108      Individuals generally experience at least one traumatic event throughout the course of

109  their lifetime (Breslau, Peterson, Poisson, Schultz, & Lucia, 2004; Copeland, Keeler, Angold,

110  & Costello, 2007; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). As such, a topic of

111  interest to many researchers is examining responses to traumatic events in order to examine

112  variability in response and clinical implications. For example, some individuals will develop

113  clinical disorders such as Posttrauamtic Stress Disorder (PTSD; Hamblen, Barnett, & Norris,

114  2012; Williams, McDevitt-Murphy, Fields, Weathers, & Flood, 2011).

115      According to Frankl (1984), meaning in life occurs when an individual accomplishes

116  personal goals, has positive interactions with others, or experiences meaningful interactions

117  with art and/or nature. What, then, is meaning in life and why is it important? Ryff &

118  Singer (1998) consider meaning to be the dedication of time to achieving personal goals of

119  importance. Battista & Almond (1973) consider meaning in life to be related to a sense of

120  "coherence." Other definitions, such as that of Crumbaugh & Maholick (1964), define

121  meaning in life as "the ontological significance of life from the point of view of the

122  experiencing individual." As evidenced by the numerous definitions presented, a coherent,

123  collectivistic definition of meaning in life seems to be lacking. This is not to say, however,

124  that the construct of meaning is unimportant (Steger et al., 2006). Naturally, the variety of

125  definitions lead researchers to measure the construct in different ways.

126      Researchers have examined the construct of meaning in life in various ways, as

127  evidenced by opposing definitions mentioned above and the great number of scales factor

128  analyzed in this study. Typically, questionnaires of this nature rely on a subjective

129  participant understanding of the construct. The Meaning in Life Questionnaire (MLQ

¹³⁰ citation), for example, asks various questions regarding participant perception of whether or

¹³¹ not their life has meaning. As an example, one question on the MLQ posits the following

¹³² statement: "I understand my life's meaning." Although this type of question seems

¹³³ transparent, a participant/client who does not have an understanding of a basic definition(s)

¹³⁴ of meaning in life will struggle to correctly interpret the construct. As a result, meaning in

¹³⁵ life measures have been criticized for their lack of objectivity in soliciting participant/client

¹³⁶ responses (Dyck, 1987; Frazier, Oishi, & Steger, 2003; Garfield, 1973; Klinger, 1977; Yalom,

¹³⁷ 1980). Coming to conclusions from existing meaning in life literature, therefore, can present

¹³⁸ a difficult challenge to researchers (Cosco et al., 2017). Additionally, meaning-based

¹³⁹ measures tend to highly correlate with other empirically-supported constructs. For example,

¹⁴⁰ Clark & Watson (1995) found that examined items relating to suicide correlated with

¹⁴¹ neuroticism, which may provide evidence for the lack of uniqueness of questions on meaning

¹⁴² in life scales. Unfortunately, however, this stems from an inability to answer fundamental

¹⁴³ questions regarding a collective definition of meaning in life (Steger et al., 2006). Further,

¹⁴⁴ Steger and colleagues argue that more effective measurements and empirically validated will

¹⁴⁵ better help researchers to understand such constructs, specifically, meaning in life.

¹⁴⁶    In examining factor structures of meaning in life measures, Chamberlain & Zika (1988)

¹⁴⁷ analyzed the factor structures of three meaning in life measures: the Purpose in Life test

¹⁴⁸ (PIL), the Life Regard Index (LRI), and the Sense of Coherence Scale (SOC). Specifically,

¹⁴⁹ this study recruited women with young children in an attempt to examine factor structures

¹⁵⁰ of the three aforementioned measures. Analyses did not suggest an underlying general factor.

¹⁵¹ However, an important limitation in this study that needs to be addressed is possible

¹⁵² differences across groups in measuring meaning in life (might go into more detail on the

¹⁵³ results from factor structure and how they contribute to why we are doing what we're doing

¹⁵⁴ with randomization). Chamberlain & Zika (1988) claim that meaning in life may be different

¹⁵⁵ for women. Then, they examine different studies conducted that have looked at gender-based

¹⁵⁶ differences (PIL; Meier and Edwards, 1974; Reker and Cousins, 1979; Shape and Viney,

157 1973). It is necessary to further examine group-based differences to identify poor items in
158 meaning in life measures.

**Scale Randomization**

160        Measurement invariance posits that different modalities of administration in different
161 settings can lead to comparable responses, which is desirable in order to negate confounding
162 of variables (Brown, 2006). Although a multitude of studies have been conducted to examine
163 administration modalities and their effects on responses (T. Buchanan et al., 2006; Meyerson
164 & Tryon, 2003; Ruyter & Wetzels, 2006), an area of concern resulting from these different
165 administration modalities is their effect on factor structure. According to T. Buchanan et al.
166 (2005), factor structure is not replicable in online vs. mailed surveys. As a result of these
167 replicability issues, researchers tend to randomize question orders in order to control for item
168 order effects, as randomizing the item order can counterbalance the scales. This ultimately
169 controls for order effects (Keppel & Wickens, 2004). However, despite controlling for order
170 effects, randomizing items on some measures could lead to changes in scale structure, as
171 some items (suicide questions or death questions) could possibly lead to different answers on
172 subsequent items. Our study sought to determine whether or not the solution of
173 randomizing items to combat the issue of replicability is problematic in and of itself.
174 Meaning in life measures, as discussed above, have the potential to lead to subjective
175 participant responses. Thus, this construct was of interest to the present study. Additionally,
176 these measures tend to be highly correlated with other constructs. For example, as discussed
177 previously, questions related to suicide tend to correlate with neuroticism. These questions
178 may lead to different-than-normal responses on subsequent questions, and this "item
179 reactivity" concept is currently being explored in a separate project (Buchanan et al, 2017).
180 However, an area of concern, aside from differing item responses, is factor structure. Thus,
181 this study sought to examine the effects of scale randomization on across a wide variety of

182  meaning in life measures at the factor level.

**Goals of Present Study**

184      Our goal for this study is to conduct a MGCFA on a wide variety of meaning in life

185  scales to examine measurement invariance across these scales. We categorized each measure

186  based on model fit by means of fit indices (cite the fit indices used here). Additionally, we

187  hoped to present noninvariant questions for each scale. As stated above, a more collectivistic

188  view of meaning in life measures is needed in regards to scale randomization (Put in more

189  information here on how we categorized these scales and how we presented noninvariant

190  questions).

**Method**

**Participants**

193      Participants in this study included 2377 students at a large Midwestern university.

194  Participants included 924 males (39.81%) and 1397 females (60.19%) between the ages of 15

195  and 55 ($M = 19.57$, $SD = 3.20$). The study included multiple different ethnicities, made up

196  of Asian participants (2.67%), Black participants (6.73%), multiple ethnicity participants

197  (2.40%), other ethnicity participants (4.94%), and White participants (83.26%). The study

198  also consisted of Freshmen (68.08%), Sophomores (18.86%), Juniors (9.11%), Seniors

199  (3.57%), and Graduate/Other students (0.39%).

**Materials**

201      need someone to make a table of the scales we used

<sub>202</sub> **Procedure**

<sub>203</sub> **Data analysis**

<sub>204</sub> **Data Screening.**

<sub>205</sub> **MGCFA.** Multigroup Confirmatory Factor Analysis (MG-CFA) was conducted on

<sub>206</sub> individual meaning in life scales. This particular process involves applying CFA principles to

<sub>207</sub> multiple groups across different each individual scale. Delivery type (non-random vs. random)

<sub>208</sub> was used to examine model fit and whether or not randomization of scales produces a worse

<sub>209</sub> or better-fitting model. We utilized previously published standards for adding restrictions to

<sub>210</sub> each MG-CFA. This approach allowed us to first examine model fit across all groups.

<sub>211</sub> Subsequently, model fit across non-random and random groups was examined. Then,

<sub>212</sub> parameters were constrained in order to calculate different types of invariances.

<sub>213</sub> *Individual Groups.*

<sub>214</sub> Utilizing a stepwise approach allowed us to examine model fit across individual groups

<sub>215</sub> by means of MG-CFA. We conducted single-group solutions based on delivery method

<sub>216</sub> (non-random question order vs. random question order). Questions delivered on paper were

<sub>217</sub> excluded for final analysis in R, as they were not part of this particular analysis. Each group

<sub>218</sub> provided us with a set of fit indices by which to evaluate model fit and examine whether or

<sub>219</sub> not scale randomization impacts factor structure across each different scale. Randomized

<sub>220</sub> scales not adhering to the published factor structure should warrant caution among

<sub>221</sub> researchers planning to deliver questions in a random format. Randomized scales adhering to

<sub>222</sub> published factor structure do not suggest any reason to avoid randomization of questions

<sub>223</sub> (Brown citation). Regardless of fit, we continued with the suggested stepwise approach by

<sub>224</sub> calculating different types of invariances. Each level of invariance adds restrictions to the

<sub>225</sub> model.

²²⁶      ***Configural Invariance.***

²²⁷      Regardless of whether or not our individualized groups both showed adequate model fit,

²²⁸ we progressed to calculate configural invariance. Configural invariance can also be referred to

²²⁹ as "equal form." This test allows the researcher to understand whether or not factor

²³⁰ structure and loadings are identical across groups, in this case non-random questionnaires

²³¹ vs. random questionnaires. This test utilizes the same set of fit indices explained above

²³² (assuming we will add this section in the data analysis section/insert a citatio).

²³³      ***Metric Invariance.***

²³⁴      Regardless of whether or not equal forms was supported across groups, we then

²³⁵ analyzed the data using metric invariance. Metric invariance examines factor loadings across

²³⁶ groups. This analysis was supported if this test of invariance did not differ significantly from

²³⁷ configural invariance. In order to meet this assumption, $\Delta$CFI $< .01$.
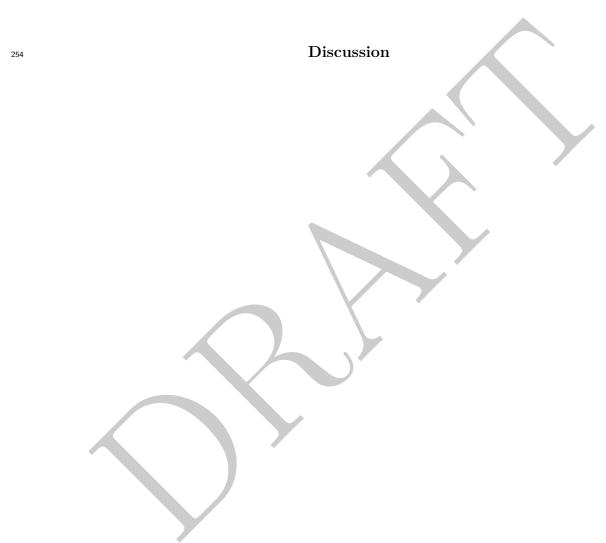
²³⁸      ***Scalar Invariance.***

²³⁹      Assuming that metric invariance did provide a large enough decrease in CFI, we then

²⁴⁰ tested scalar invariance. Scalar invariance examines indicator intercepts and determines

²⁴¹ whether or not these are equal across groups. Additionally, scalar invariance determines

²⁴² whether or not group membership influences a role in raw scores across groups. If the change

²⁴³ in CFI is not equal to or greater than .01, this assumption has been met. As with metric

²⁴⁴ invariance, this analysis was supported if the test of invariance did not differ significantly

²⁴⁵ from configural invariance.

²⁴⁶      ***Partial Invariance.***

²⁴⁷      Different methods have been utilized for scales that differ when utilizing the stepwise

²⁴⁸ method for conducting the different types of invariances. The scale can either be abandoned

²⁴⁹ or the noninvariant items removed for further analyses. This may affect construct validity as

250 well as the theory behind the scale (Cheung & Rensvold, 1999). We relaxes constructs of

251 noninvariant items for the remainder of analysis, as suggested by Brown (2006) & Byrne et

252 al. (1989).


253                                     **Results**


254                                    **Discussion**

<sub>255</sub> **References**

<sub>256</sub> Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology:*

<sub>257</sub> *Research and Practice*, *33*(2), 148–154. doi:10.1037/0735-7028.33.2.148