Cultural Adaptation of Interventions and the *a Priori* Assessment of Intervention Fit: Exploring Measurement Invariance for American and Swedish Youth Leaving Care

Russell Turner, Emily G. Vira, Martin Bergström & Tina M Olsson

Published online: 28 Dec 2022.

Submit your article to this journal ⬈

Article views: 173

View related articles ⬈

View Crossmark data ⬈

Routledge
Taylor & Francis Group

Cultural Adaptation of Interventions and the *a Priori* Assessment of Intervention Fit: Exploring Measurement Invariance for American and Swedish Youth Leaving Care

Russell Turner[a], Emily G. Vira[b,c], Martin Bergström[d], and Tina M Olsson[a,e]

[a]Department of Social Work, University of Gothenburg, Gothenburg, Sweden; [b]Department of Learning, Informatics Management and Ethics, Karolinska Institute, Stockholm, Sweden; [c]Department of Mental Health and Suicide, Norwegian Institute of Public, Oslo, Norway; [d]School of Social Work, Lund University, Lund, Sweden; [e]School of Health and Welfare, Jönköping University, Jönköping, Sweden

ABSTRACT

Purpose: There is a growing literature on the importance of cultural adaptation of research-supported social work interventions. Few studies have however offered systematic methods for the *a priori* assessment of intervention fit in a new context. The current study explores the use of measurement invariance analyses to help identify whether key theoretical constructs in an intervention's theory of change may fit differently in a new context.
Methods: We draw on data on 13 measures of key constructs in an intervention for youth leaving out-of-home care designed and trialed in the US context (N = 295; 53.1% girls, mean age 17.3) compared to Swedish adolescents (N = 104; 41% girls, mean age 17.5).
Results: In general, the results found all the measures to be invariant between US and Swedish samples.
Discussion: The original intervention (US) is likely to have a good fit in the new (Swedish) context in terms of the measurement and functionality of its key constructs. There are, however, some indications that certain aspects of the original key constructs may function differently in the Swedish context, highlighting a need to review either measurement or intervention design. A broader conclusion is that measurement invariance can provide a useful tool for research-supported social work practice, namely the systematic *a priori* assessment of the transferability of an intervention's theory of change. Some limitations and methodological issues are discussed.

A key challenge for research-supported social work is the transfer of successful interventions to new contexts (Castro et al., 2006; Sundell et al., 2016). Prior to implementing an intervention in a new context, there needs to be an evaluation of the potential fit of the intervention, such as the ability to achieve similar results to those found in the original context (Von Thiele Schwarz et al., 2019). This raises the broader question of how the *a priori* assessment of intervention fit for the new context may be systematically undertaken. Specifically, there is a need to assess the fit of the measures, which were used in the original trial, in the new context (Ferrer-Wreder et al., 2012). Youth transitioning from out-of-home care is a group where different interventions are in use, but few have been systematically

CONTACT Russell Turner russell.turner@socwork.gu.se Department of Social Work, University of Gothenburg, Box 720, 405 30 Gothenburg, Sweden

assessed in terms of the fit of intervention measures in the new implementation setting (see, Bergström et al., 2020). If the original measures perform differently in the new setting, then conclusions about implementation may not be correct. Moreover, the intervention's key constructs may be understood or function differently in the new context, and thus practice based on these constructs may need reviewing prior to implementation. In this paper, we explore whether an *a priori* assessment, by means of measurement invariance testing, can help assess the fit of a social work research-supported intervention in a new context. To do this, we use a case study of an intervention for youth leaving out-of-home care, trialed in the US, and assess its fit in a new cultural context (Sweden).

Models for intervention transfer and fit

Broadly, intervention fit can be thought of as the suitability of an intervention in a new setting. There are several different models, however, of conceptualizing intervention transfer and fit, with most models advocating a balance between fidelity to the original theory of change *and* adaptation to the local context (Castro et al., 2006; Von Thiele Schwarz et al., 2019). Empirical work on intervention transfer has also drawn attention to the importance of assessing fit in the new context for an intervention's effectiveness, with either "home-grown" or contextually adapted interventions performing better than interventions that are adopted wholesale (Sundell et al., 2016). For interventions to be transferable between contexts, however, there needs to be some *a priori* measure of the approximate cultural or contextual fit. Moreover, that such an assessment takes a systematic approach. Bauman et al. (2015) stated that there are two main methods to how interventions should or can be adapted: those that focus on the content of the intervention and those that inform the process of adaptation. In both approaches, an important first step is some *a priori* assessment of fit in the new context. There are, however, few systematic tools by which to do this.

Assessing intervention fit in a new context

A fruitful way of assessing intervention fit in a new context may be found in using the interventions' theory of change (ToC) as a set of fit criteria. Interventions are usually built on the idea of changing one or several key constructs, which are usually explicated in the intervention's ToC. For example, in the case of programs to support youth transitioning from out-of-home care, interventions focus on key constructs such as building resilience and self-efficacy. In evaluations of interventions for youth transitioning from out-of-home care, a battery of measures is used that assess both processes, such as participation in particular competence-building activities, as well as proximal outcomes, such as changes in resilience and self-efficacy (see, Olsson et al., 2020). Some studies have also used such measures to prospectively explore the appropriateness of an intervention for trial in a new cultural context (Olsson et al., 2020). Thus, an intervention's evaluative measures, if sufficient in scope to cover key constructs in the intervention's ToC, may also be used to assess *a priori* the intervention's cultural fit in a new context.

One such tool to assess the intervention's fit using evaluative measures, is equivalence or measurement invariance testing. There is an emerging research corpus which has used measurement invariance analysis to systematically assess differences in the fit of measures between cultural settings (Boer et al., 2018; Jovanovic et al., 2019). Measurement invariance

studies are primarily concerned with the retrospective assessment of whether sub-groups, in terms of culture, gender, and so on, respond differently. This is to ensure comparability – invariance/equivalence of measures – in order to pool data, make comparisons, and draw conclusions across different settings and groups. Whilst this is an important function of measurement invariance analysis, we wish in the current article to explore the role of measurement invariance analysis for a broader discussion of the *a priori* assessment of cross-cultural or cross-setting fit of interventions.

Measurement invariance and intervention fit

Measurement invariance studies are often used because people in different cultural contexts may attach different meanings to items, e.g., on a questionnaire (Byrne, 2016; Chen, 2008). Thus, a universalist way of understanding measurement invariance as a tool is that the key constructs in the intervention's ToC should perform similarly across cultural groups. A contextualist interpretation, which we use in the current article, is that measurement invariance analysis can be used to assess whether key constructs from one context are functionally equivalent in another cultural context. If measures have a different fit, this alerts us to potential transferability issues, first in terms of the measurements of the core intervention components. Poor fit of measures could also highlight aspects of key constructs in the ToC that may function differently in the new context. In other words, researchers could assess potential cultural fit using the intervention's measures prior to the intervention being imported and trialed.

Put simply, measurement invariance testing is a series of sequential tests that assess whether independent groups have an equivalent or non-equivalent interpretation of the same measure (Putnick & Bornstein, 2016). Measurement invariance testing can be implemented by conducting multiple group confirmatory factor analyses (MGCFA), with increasing model constraints to test three different levels of invariance (Millsap, 2012). The first level is configural invariance, which requires the factor structures to be equivalent across groups. A poor-fitting configural model would indicate that the key concept, e.g., resilience, may not be comparable between groups in the form that it is being measured. If configural invariance is not assumed no further equivalence tests can be meaningfully assessed (Vandenberg & Lance, 2000).

If configural invariance is assumed, the next level of invariance is metric invariance, which, in short, assesses whether the participants in both groups equally understand the items' meaning in relation to the measured construct of interest (see, Putnick & Bornstein, 2016). If metric invariance is assumed, it is inferred that the participants from both groups give equal meaning to the items in relation to the core construct being measured. For example, an item from the Resilience scale is "I can get through difficult times because I have experienced difficulty before" (Wagnild & Young, 1993). Metric invariance of the item would mean that both groups respond to this item equally as an indicator of resilience. If the opposite is true, and the items' factor loading is non-invariant or non-equivalent, this would mean that the item may be more related to resilience in one group than the other, i.e., indicative of differing cultural interpretations regarding resilience.

The final level of invariance is scalar invariance, in which the intercepts (the means of each item) are constrained to be equal across groups. This assesses whether the participants in each group have sufficiently similar means on a given item. This is particularly important
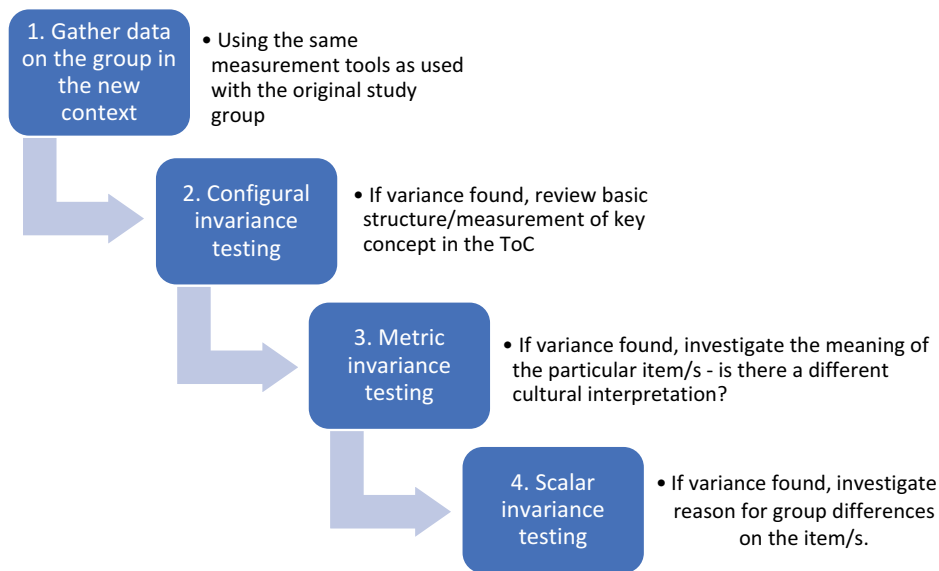
Figure 1. Flow-chart of the measurement invariance testing process.

for comparing latent construct mean differences between groups. For example, the following item "I feel proud that I have accomplished things in life" from the Resilience scale (Wagnild & Young, 1993) may lead to different mean points between two different cultures. In American culture, showing immense pride for achievements may be more socially desirable than in a comparison culture, that may either be more collectivistic or conservative (Furukawa et al., 2012). If mean levels of pride of accomplishments are too dissimilar, i. e., scalar non-invariance, then this suggests the measure, if not any intervention components related to a personal sense of achievement, may need re-visiting. If a measure is found to be invariant – at metric or scalar level – a partially invariant model can also be tested releasing the constraint of the non-invariant item and re-assessing whether model fit is not substantially worse than the previous configural model (Millsap & Kwok, 2004; Putnick & Bornstein, 2016). Figure 1 depicts an overview of the measurement invariance testing process.

The current study

The current study aims to address whether measurement invariance analyses help identify whether key theoretical constructs, in an intervention's ToC, fit differently in a different cultural context. By "key theoretical constructs" we refer to the intervention's "basic integrity" (see, Sundell et al., 2016), that is, the central features of the ToC. Whilst on the one hand these are theoretical constructs, as applied constructs they are practical, embodied, material, or technical aspects of intervention delivery. For example, in terms of interventions for youth transitioning from out-of-home care, key constructs relate to things like promotion of independence. Whilst this has an abstract theoretical element, in terms of intervention delivery it relates to practical process outcomes that youth are expected to have participated in, such as opening a bank account. We address our aim using a case study of an intervention for youth leaving out-of-home care (see, Blakeslee et al., 2020). The

intervention, based on self-determination theory, was designed and trialed in the US context. The measures applied in the US intervention were used to assess a Swedish sample of youth transitioning from out-of-home care to independent living (see, Olsson et al., 2020). Our main research question is: To what extent do measurement invariance analyses help identify areas of key theoretical constructs in the intervention's ToC that may fit differently in different settings?

We assess measurement invariance of the process and proximal outcome measures used in both settings to explore what aspects of the intervention might have been amenable to *a priori* adaptation. To our knowledge, a cultural comparison involving measurement invariance of the outcome measures used in the My Life self-determination model of intervention has not been previously investigated. Thus, a secondary aim with our study is to report on the transferability of the My Life self-determination model of intervention from a US to a Swedish context, with the research question: What level of invariance – out of configural, metric, and scalar – is achieved for each scale when comparing the US and Swedish samples?

Method

The current study uses two separate archival cross-sectional data sets on interventions for youth transitioning from out-of-home care. The first dataset is from the US (My Life programs) and was collected between 2010 and 2013 in the greater Portland, Oregon metropolitan area as part of two concurrent, large-scale randomized trials of the *My Life* model (Blakeslee et al., 2020). The trials involved adolescents in foster care ($N = 295$; 53.1% girls, 46.9% boys; mean age 17.3, SD = .61). Only the baseline data is used in the current study. The second dataset is on Swedish adolescents (N = 104; 41% girls, 59% boys; mean age 17.5, SD = .14) placed in out-of-home care from a previously published study (Olsson et al., 2020). The Swedish data were collected between February and October 2019. For more detail on recruitment for both samples, see, Olsson et al. (2020).

Measures

We analyze five measures that represent either process or proximal outcomes of the intervention. The use of a range of measures is to capture aspects of the program's core components. Each of the five measures is briefly described below. Further information about reliability and missing data is available in Olsson et al. (2020).

My Life Activity Checklist (MLAC)

The MLAC is a 44-item (yes/no) checklist designed to assess the extent to which youth have participated in or performed certain activities important for preparing to live independently (Wehmeyer & Palmer, 2003). The measure has three subdomains: career development (e.g., created a CV, talked to a career advisor); postsecondary education development (e.g., talked to a teacher or guidance counselor about going to college, got information about financial aid or scholarships to pay for college); and daily life activities (e.g., opened a bank account, scheduled an appointment with a case manager or professional in the community). Three free-text items were removed from the checklist for the current study as were two further

items that were not applicable to the Swedish welfare context ("Got my social security card," "Applied for health insurance").

Youth Transition Planning Assessment (YTPA)

The 17-item YTPA (Powers et al., 2001) measures transition planning knowledge and engagement using a 4-point ordinal scale. The measure includes items such as, "People ask about my opinions and ideas at meetings" and "I help run my transition planning meetings." One item (item 17) was removed as its response items were categorical.

My Life Self Efficacy Scale (MLSES)

The 17-item *My Life* Self-Efficacy Scale (Blakeslee et al., 2020) assesses youths' beliefs about carrying out the skills targeted by the *My Life* intervention (e.g., problem solving, self-monitoring, working with adults) on a 5-point ordinal Likert scale. It includes three subscales: Self-regulation (e.g., I am confident that I can keep myself from being over-whelmed by stressful situations); Managing others (e.g., I am confident that I can make agreements with adults to help me in specific ways); and Achievement (e.g., I am confident that I can solve problems that keep me from achieving goals).

Career Decision-Making Self-Efficacy – short form (CDSE)

The 25-item CDSE assesses the degree of belief an individual has that they can successfully complete tasks necessary to making career decisions (Betz et al., 1996). The measure includes the following subscales: Accurate Self-appraisal, Gathering Occupational Information, Goal Selection, Making Plans for the Future, and Problem-solving. Item examples include, "Determine what your ideal job would be"; "Change majors if you did not like your first choice"; and "Successfully manage the job interview process."

Resilience Scale (RS)

The resilience scale assesses levels of resilience in the general population (Wagnild & Young, 1993). The short version RS-14 was adopted in the current study and includes items such as "I can get through difficult times because I've experienced difficulty before" and "My belief in myself gets me through hard times."

Handling of missing data

The two separate samples were investigated for missing values. For most measures, missing data on single items were acceptable (less than 5%) and thus data were missing completely at random (MCAR; Little's MCAR test, $p > .05$). However, for a number of measures, the Swedish sample had missing values on single items exceeding 5%. In addition, for two measures in particular (Youth Transition Planning Assessment, My Life Self-Efficacy Scale) it was not possible to determine whether data was MCAR. Therefore, multiple imputation was used to impute missing values on single items across all measures for each sample individually.

Analytical plan

We conducted all statistical analyses in R Statistical Software (version 4.0.3, R Core Team, 2020), using the packages "tidyverse," "rio," "lavaan" (Rosseel, 2012), and "MVN." Visual inspections of the data and Mardia tests showed that data (on the ordinal measures) were skewed. Several authors have suggested methods for dealing with such non-normal data (e. g., Finney & DiStefano, 2006; Rhemtulla et al., 2012). Following these authors, we used a maximum likelihood estimator with robust standard errors (MLR) and estimates calculated using a Yuan-Bentler correction. Previous work has demonstrated that such a correction produces unbiased estimates even given non-normal data (Lai, 2019).

Measurement invariance was analyzed using MGCFA. Due to low power from the relatively small sample size, we tested measurement invariance on the sub-scales as first-order factors. To assess for measurement invariance between groups (Swedish vs. US), three models were assessed: configural, metric, and scalar. A configural model contained unconstrained parameters, a metric model contained constrained factor loadings, and a scalar model contained constrained factor loadings and intercepts. The following goodness-of-fit determinants, as suggested by Little (2013), were used to assess configural model fit: Chi-square (χ^2), robust Comparative Fit Index (CFI) > .90, robust root mean square error of approximation (RMSEA) < .08, robust RMSEA confidence intervals, and robust Tucker–Lewis index (TLI) > .90. Measurement invariance was primarily determined using a change of >.01 in CFI (ΔCFI; Cheung & Rensvold, 2002). We also, however, assessed invariance using statistically significant change in chi-square ($\Delta\chi2$) as a secondary measure as using $\Delta\chi2$ alone has been associated with finding higher levels of metric and scalar invariance (Putnick & Bornstein, 2016). If non-invariance was observed, a lavaan test score function (see, Rosseel, 2012) was used to identify the main source of non-invariance. A subsequent partial invariant model was assessed with freed non-invariant item(s). Partial invariant models with released parameters have been shown to lead to un-biased estimates when comparing groups, where the researchers can be certain that the same construct is being measured (Shi et al., 2019).

Results

The presentation of the results begins with an overview of the levels of measurement invariance achieved for each scale or sub-scale. This is followed by sub-sections that presents more detailed results for each scale. Table 1 shows the scales and sub-scales and the level of measurement invariance achieved. Thirteen scales or sub-scales were tested. Eleven achieved configural invariance, meaning that the key concept is comparable across contexts. The models for two scales did not converge. Ten scales achieved metric invariance meaning both groups give equal meaning to the items that measure the core construct. Two scales, however, achieved only partial metric invariance, meaning non-comparability of meaning on specific items. Nine scales achieved scalar invariance, which highlights similarity of means on given items. Six of these scales achieved only partial scalar invariance, meaning dissimilarity on specific items.

Table 1. Overview of measurement invariance results by scale.

Measures and subscales	Configural	Metric	Partial Metric	Scalar	Partial Scalar
MLSES (Table 2)					
Self-regulation	Yes	Yes	–	Yes	–
Managing Others	Yes	Yes	–	Yes	–
Achievement	Yes	Yes	–	No	Yes
MLAC (Table 3)					
Career	N/A	N/A	N/A	N/A	N/A
Education	Yes	Yes	–	No	No
Daily Life	N/A	N/A	N/A	N/A	N/A
YTPA (Table 4)	Yes	No	Yes	–	–
CDSE (Table 5)					
Self-appraisal	Yes	Yes	–	No	Yes
Gathering Occupational Information	Yes	No	Yes	No	Yes
Goal Selection	Yes	Yes	–	Yes	–
Problem-solving	Yes	Yes	–	No	Yes
Making Plans	Yes	Yes	–	No	Yes
Resilience	Yes	No	Yes	No	Yes

Table 2. Measurement invariance of the my life activity checklist sub-scales.

Structural models	Robust X^2	df	p	Δ X^2	Δdf(p)	Robust RMSEA	95% CI RMSEA	TLI	CFI	ΔCFI
Career subscale										
Configural	(model did not converge)									
Education subscale										
Configural	150.46	88	<.01	-	-	.061	.044;.077	.969	.975	
Metric	168.75	98	<.01	16.63	10(.08)	.061	.045;.077	.968	.972	−.003
Scalar	159.89	97	<.01	**16.27**	**1 (<.01)**	.058	.041;.074	.971	.975	+.003
Daily life subscale										
Configural	(model did not converge)									

Note: Significant differences ($p < .05$ and ΔCFI > .01) are shown in bold.

My Life Activity Checklist (MLAC)

Table 2 shows the measurement invariance results for the MLAC subscales of Career, Education, and Daily Life Activities. The configural model for the Career sub-scale did not arrive at an acceptable solution. The Education sub-scale showed good configural model fit and was also invariant at the metric level. Scalar invariance was found (according to ΔCFI but not $Δχ^2$). The lavaan score test showed that items 16 "Visited a college or vocational school" and 22 "Completed an application for college or vocational school" were non-invariant. The Swedish sample had higher (more positive) means on item 16, while the US sample had higher means on item 22. A partial scalar model was tested by releasing the items. This model, however, was still non-invariant. Invariance testing stopped at this point as model convergence problems arose on attempting to release further items. The configural model for the sub-scale Daily Life Activities did not converge.

Youth Transition Planning Assessment (YTPA)

Table 3 shows the measurement invariance tests for the YTPA. The configural model showed less than adequate fit. Despite this, metric invariance was tested and non-invariance was found. The lavaan test score revealed three non-invariant items: item 7 "I know how to get the help I need when making plans for my future," item 2 "I understand how school can

Table 3. Measurement invariance of the youth transition planning and assessment scale.

Structural models	Robust X^2	df	p	ΔX^2	Δdf(p)	Robust RMSEA	95% CI RMSEA	RobustTLI	RobustCFI	ΔCFI
Total scale										
Configural	850.96	208	<.01	-	-	.137	.128;.147	.742	.777	-
Metric	916.53	223	<.01	65.82	**15(<.01)**	.137	.128;.146	.742	.760	**.017**
Partial Metric	873.41	220	<.01	21.60	**12(.043)**	.125	.117;.133	.753	.774	.003

Note: Significant differences (*p* < .05 and ΔCFI > .01) are shown in bold.

help me plan for the future," and item 3 "I understand how DHS can help me plan for the future." The Swedish sample had higher (more positive) loadings on all three of the items. A partial metric model was tested, which resulted in partial metric invariance (Δχ² was borderline significant but ΔCFI was not).

My Life Self-Efficacy Scale (MLSES)

Table 4 shows the measurement invariance results comparing the US My life sample to the Swedish sample for the subscales of Self-regulation, Managing Others, and Achievement, from the MLSES scale. The Self-regulation subscale and the Managing Others subscale both showed good model fit and assumed measurement invariance between US and Swedish samples, on all three levels of invariance. The Achievement subscale showed good model fit on the configural and metric level but not on the scalar level. The lavaan score test indicated that items 1 "I am confident that I can describe in detail my dreams for adult life" and 17 "I believe I can accomplish the goals set for myself," were non-invariant. The Swedish sample had a higher mean on the first item, whereas the US sample had a higher mean on the second item. A subsequent partial scalar showed good model fit.

Career Decision-making Self-efficacy scale (CDSE)

Table 5 shows the measurement invariance results for the subscales of Self-appraisal, Managing Others, Gathering Occupational Information, Goal Selection, Problem Solving, and Making Plans from the CDSE scale. The Self-appraisal subscale showed good configural model fit. Metric invariance was found, but not scalar invariance. The lavaan score test indicated that item 5 "Accurately identify what is easy and hard for you to do" was non-invariant. The US sample had a higher mean on the particular item. A subsequent partial scalar model showed perfect fit and partial scalar invariance was assumed. The Gathering Occupational Information subscale showed good configural model fit. Metric non-invariance was found and the lavaan score test indicated that item 10 "Find out what the future is for a career area over the next ten years" and item 15 "Find out how much people make in a particular career" were metrically non-invariant. The US sample had a higher mean. A subsequent partial metric showed good model fit. Lastly, scalar invariance was achieved (with the freed factor loadings). The Goal Selection subscale showed good model fit and was invariant at all three levels. The Problem-Solving subscale showed good model fit and was metrically invariant but showed scalar non-invariance. The lavaan score test indicated that item 8 "Continue to work on your career goal even when you get frustrated" was non-invariant. The US sample had a higher mean. The Making Plans subscale showed good configural model fit and was invariant at the metric, but not scalar level. The lavaan score

Table 4. Measurement invariance of the my life self-efficacy sub-scales.

Structural models	Robust X^2	df	p	Δ X^2	Δdf(p)	Robust RMSEA	95% CI RMSEA	Robust TLI	Robust CFI	ΔCFI
Self-regulation subscale										
Configural	5.58	4	.233	-	-	.050	.000;.137	.982	.994	-
Metric	9.54	7	.216	3.95	3(.27)	.047	.000;.113	.984	.991	.003
Scalar	14.83	10	.138	5.44	3(.14)	.053	.000;.106	.980	.983	.008
Managing Others subscale Configural	27.77	18	.066	-	-	.058	.000;.098	.972	.982	-
Metric	30.83	23	.127	2.35	5(.80)	.045	.000;.082	.983	.987	−.005
Scalar	38.25	28	.094	7.53	5(.18)	.046	.000;.079	.983	.984	.003
Achievement subscale Configural	31.89	18	.023	-	-	.069	.026;.107	.957	.974	-
Metric	32.19	23	.096	1.34	5(.93)	.050	.000;.088	.977	.982	−.008
Scalar	63.29	28	.000	35.74	**5(.00)**	.088	.059;.117	.929	.934	**.048**
Partial Scalar	34.76	26	.117	2.19	3(.53)	.046	.000;.082	.981	.983	−.001

Note: Significant differences (p < .05 and ΔCFI > .01) are shown in bold.

test indicated that item 7 "Decide the steps you need to take to successfully complete your classes" was non-invariant. The US sample had a higher mean. A subsequent partial scalar model showed adequate model fit.

In sum, the subscales of the CDSE scale showed mixed fit and a few non-invariant items were identified. The Self-appraisal subscale, Making Plans subscale, and Problem-Solving subscale achieved partial scalar invariance. The Gathering Occupational Information subscale had partial metric invariance but then assumed scalar invariance with two freed factor loadings. Lastly, the Goal Selection subscale achieved invariance between the US and Swedish samples on all levels of measurement invariance.

Resilience Scale (RS-14)

Table 6 shows the measurement invariance results for the RS-14. The RS-14 showed good configural model fit. However, metric invariance was not found. The lavaan score tested indicated that item 3 "I usually take things in stride" was non-invariant. The Swedish sample had a higher loading than the US sample. A partial metric model with the freed factor loading showed good model fit, although scalar invariance was not found. The lavaan score test indicated that item 2 "I feel proud that I have accomplished things in life," item 1 "I usually manage one way or the other," item 3, and item 13 "My life has meaning" were non-invariant. The US sample had a higher mean on each of the particular items. A subsequent partial scalar model with the freed intercepts showed good model fit and partial scalar invariance was achieved.

Discussion

There is a growing literature in the field of research-supported social work practice that argues for or examines the importance of cultural adaptation of interventions (e.g., Bauman et al., 2015; Castro et al., 2006; Sundell et al., 2016). Bauman et al. (2015) maintained that an important first step in considering cultural adaptation is the careful analysis of the new target population and the target domain of the intervention; in short, the assessment of fit of the intervention in the new context as well as transferability of its measures. While there are

general frameworks and recommendations for assessing fit, there is a need to develop systematic tools for assessing potential transferability of interventions, measurement of key constructs, and what, if anything, would need to be adapted to the new cultural context. We connect to this literature with the aim of illustrating one such systematic method for the *a priori* assessment of intervention fit in a new context. Using pre-implementation measures from samples of American and Swedish youth leaving care, measurement invariance testing was explored as a method of systematically assessing the fit of the My Life self-determination model intervention in a Swedish context.

Five comprehensive measures that represented key constructs in the My Life self-determination model's theory of change for youth leaving out-of-home care were tested, with 11 (of 13 scales) that converged. If measures are non-invariant between contexts this firstly means that measures may not be comparable between contexts. Second – and perhaps more importantly – non-invariance suggests that key concepts in the intervention may understood differently in the new context. In general, the scales were invariant between US and Swedish samples. Thus, regarding our secondary research question, all tested scales achieved configural invariance, 10 scales achieved metric invariance, and nine achieved scalar invariance. As configural invariance was found for all scales, this would suggest that the key constructs in the original (US) intervention's ToC are likely to have a good fit in the new (Swedish) context, at least at a general level.

There were, however, some indications that certain aspects of the original key constructs may function differently in the Swedish context. This is where six of the measures only attained partial invariance. This means that the US and Swedish samples potentially construed different meanings to certain items or had different levels of some aspects of the target behavior or construct. An example of metric invariance concerned the Resilience scale. The higher loading of the item "I usually take things in my stride" for the Swedish sample suggests that this characteristic – let's call it "unflappability" – was more strongly linked to resilience in Sweden. This metric invariance could alert program deliverers to differences in the cultural expression of resilience. In other words, being resilient may look more like "unflappability" in Sweden than in a US context. Program delivery and responsivity to participants could thus be more attuned to this potential cultural difference.

An example of scalar invariance, i.e. where the US and Swedish samples had different means, concerned the Achievement subscale of the self-efficacy measure, and in particular the items "I am confident that I can describe in detail my dreams for adult life" and "I believe I can accomplish the goals set for myself." The Swedish sample tended to score higher on these items, suggesting that the intervention as originally delivered may not impact in the same way, e.g., due to ceiling effects of some items. In this particular case, this could mean that practitioners in the Swedish context may need to give less focus to these aspects of achievement self-efficacy. This does not necessarily mean that intervention's central concepts in the ToC should be changed. There is a need to be mindful of "overfitting" intervention delivery to the evaluation measures, resulting in biasing delivery toward outcome measures. On the other hand, if an intervention intends on changing a key construct, such as building self-efficacy, but components of this key construct appear already to be in place, then delivery on these specific aspects may already be reaching a saturation point. The partial invariance results we found rather support the idea of general transferability of the key construct, with some reservation for potential differential effect.

Table 5. Measurement invariance of the career decision-making self-efficacy sub-scales.

Structural models	Robust X^2	df	p	ΔX^2	Δdf (p)	Robust RMSEA	95% CI RMSEA	Robust TLI	Robust CFI	ΔCFI
Self-Appraisal subscale Configural	10.84	10	.370			.026	.000;.101	.996	.998	
Metric	12.27	14	.585	.45	4 (.98)	.000	.000;.072	1.01	1.00	−.002
Scalar	23.19	18	.184	**13.53**	**4 (.01)**	.044	.000;.089	.987	.988	**.012**
Partial Scalar	17.15	17	.445	5.71	3 (.13)	.008	.000;.075	1.00	1.00	.000
Gathering Occupational Information subscale Configural	7.79	10	.650			.000	.000;.069	1.00	1.01	
Metric	19.35	14	.152	**12.52**	**4 (.01)**	.047	.000;.093	.983	.988	**.022**
Partial Metric	9.68	11	.559	1.92	1 (.17)	.000	.000;.073	1.00	1.01	.000
Scalar	15.57	15	.411	6.27	4 (.18)	.015	.000;.073	.998	.999	.01
Goal Selection subscale Configural	15.24	10	.124			.065	.000;.126	.979	.989	
Metric	20.15	14	.126	4.15	4 (.39)	.055	.000;.105	.985	.989	
Scalar	25.69	18	.107	5.48	4 (.24)	.053	.000;.096	.986	.987	.002
Problem-Solving subscale Configural	27.22	10	.002			.093	.061;.126	.923	.962	
Metric	34.28	14	.002	5.16	4 (.27)	.104	.006;.148	.943	.960	.002
Scalar	49.11	18	.000	**16.81**	**4 (.00)**	.109	.073;.146	.937	.943	**.017**
Partial Scalar	39.12	17	.002	6.87	3 (.27)	.095	.056;.135	.952	.959	.001
Making Plans subscale Configural	41.35	10	.000			.130	.091;.173	.898	.949	
Metric	42.58	14	.000	1.74	4 (.78)	.106	.071;.143	.933	.953	−.004
Scalar	56.09	18	.000	**13.66**	**4 (.01)**	.106	.076;.138	.933	.939	**.014**
Partial Scalar	46.58	17	.000	3.52	3 (.32)	.097	.064;.131	.944	.952	.001

Note: Significant differences ($p < .05$ and ΔCFI > .01) are shown in bold.

Table 6. Measurement invariance of the short-version resilience scale.

Structural models	Robust X^2	df	p	ΔX^2	Δdf(p)	Robust RMSEA	95% CI RMSEA	Robust TLI	Robust CFI	ΔCFI
Resilience Model: Configural	312.21	154	.000			.082	.069;.095	.910	.924	
Metric	339.15	167	.000	**27.03**	**13(.01)**	.082	.069;.094	.911	.914	**.01**
Partial Metric	333.62	166	.000	20.97	12(.05)	.081	.069;.094	.913	.920	.004
Scalar	408.29	179	.000	**87.68**	**13(.00)**	.090	.079;.102	.892	.893	**.027**
Partial Scalar	350.02	175	.000	15.95	9(.07)	.080	.068;.093	.915	.918	.002

Note: Significant differences ($p < .05$ and ΔCFI > .01) are shown in bold.

Small discrepancies between cultures may well be frequent but these need not always impede intervention transferability at a general level.

The above discussion focused on intervention constructs or targets that may be *less* impacted upon in the new context. Further inspection of other measures in the current study revealed aspects of the intervention's key constructs that may need to be adapted to *increase* response in the new context. The Gathering Occupational Information subscale of the Career Decision-Making Self-efficacy measure was not scalar invariant; the item "Find out how much people make in a particular career" showed that the Swedish sample tended to score lower on this item. A potential explanation of this finding is that speaking about salaries and earnings is generally seen as a more private matter in Sweden. Practitioners might then need to re-focus any tasks related to this skill toward gather general information about salaries, e.g., from websites rather than individuals. Interestingly, the Swedish sample also had a lower mean on the item "Decide the steps you need to take to successfully complete your classes," suggesting a lower baseline knowledge. In practice, this may mean that additional or new intervention elements may be needed to work effectively with this lower baseline. These two items together may speak to cultural differences in the schooling systems between the original and new context.

Our main research question concerned whether measurement invariance analyses help identify areas of key theoretical constructs in the intervention's ToC that may fit differently in different settings. The above results illustrate how measurement invariance can be a useful tool in the assessment of cultural fit for research-supported social work interventions. Based on the findings and literature on measurement invariance, we suggest aiming for, at minimum, configural invariance (i.e. the construct is similarly understood across samples), then metric invariance (the items comprising the construct are understood similarly). Scalar invariance is also important to assess whether levels of need in the new context are sufficiently similar to expect comparable impact.

Turning now to the broader point of the *a priori* assessment of intervention fit, the systematic assessment of intervention's transferability needs to be a cost- and time-effective exercise. In the current study, we used existing baseline data (from the original US site) and cross-sectional survey data on the Swedish sample. Once the data were gathered, the measurement invariance analysis is a fairly low-cost exercise, presuming a level of familiarity with the intervention measures, as well as measurement invariance testing. While the gathering of survey data on the new, intended population requires planning and co-ordination, there are several other benefits to gathering such data, such as assessment of initial need and providing pre-implementation baseline measurement for later evaluation. Moreover, there is an increasing push toward making data sets available for use outside of the original research groups, which could aid in having comparison datasets for analysis (Gilmore et al., 2018).

There are, however, some problems that can occur with using measurement invariance as a method for the *a priori* assessment of intervention fit. One issue encountered in the current study was that two models did not converge. This was most likely due to low sample size in relation to the complexity of the model. Thus, larger sample sizes are to be recommended, even though much can be done with fairly small samples. However, use of larger samples may be difficult when assessing the possibility of transferring an intervention for very specific target groups where large samples are simply not possible. In the current study, we tested some subscales as first-order factors, rather than as second-order factors as

part of a larger model. We see this as a pragmatic way of dealing with lower sample sizes, even though we acknowledge as this may depart from a stricter psychometric understanding of how sub-factors are linked to higher order factors. Furthermore, we used the measures as though all were fully validated. Some measures had less than adequate configural fit, e.g., the YTPA scale. Rather than the poor fit being related to differences between the US and Swedish sample, the poor fit may relate to unsound psychometric properties in the original scale. The YTPA has previously been used a single factor measure by Powers et al. (2001), but they also reported that an exploratory factor analysis suggested two factors. To our knowledge, there is no prior psychometric evaluation of this measure. Prior to using measurement invariance to assess program transference possibilities, we would therefore suggest that either validated measures are used or that preparatory testing is undertaken on the intended measures. A further point here is that a measure may have good psychometric properties based on samples in the original cultural context but not necessarily in the new context, particularly if scales need to be translated into a different language. The scales used in the current study were translated and back-translated, then piloted with Swedish adolescents, which may assist the successful adaptation of measures (Bornman et al., 2010). Another issue is that not only were the sample sizes small, but the samples had unequal group sizes which may bias MGCFA results (Yoon & Lai, 2018). Lastly, we chose to assess change in model fit prioritizing ΔCFI over $\Delta\chi^2$, as the Chi-square statistic has been shown to be sensitive to non-normality and over-reject well-fitting models (Andreassen et al., 2006; Powell & Schafer, 2001).

A broader issue with using measurement invariance as a tool for assessing fit for social work interventions is that the assessment will only be as good as the match between the measures used and intervention's core constructs. In other words, the assessment using measure invariance needs to use a broad range of measures that captures different, but central aspects of the intervention's theory of change. It should also be noted that there are likely to be aspects of an intervention's cultural fit that are harder to assess using psychometric measures. Using measurement invariance may provide a systematic tool for assessing intervention fit, but it needs to be located in a broader model of understanding cultural adaptation of interventions within social work practice (see, Bauman et al., 2015).

In summary, the current study has provided an argument for and an empirical example of the use of measurement invariance for the *a priori* assessment of an intervention's fit in a new context. The results demonstrated the possibility of using such a tool, as well as highlighting how specific adaptations to an intervention's delivery, which may be necessary in the new context, can be discerned from the measurement invariance analyses. Going forward, we identified several issues that researchers may wish to consider when using measurement invariance as a tool for systematically assessing an intervention's fit in a new cultural context.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

The study is based on data collected for a previous study that was granted ethical approval on December 3rd 2018 by the Swedish regional ethics board (Dnr. 742-18). Informed consent was obtained from all individual participants included in the study. The current study was performed in line with the principles of the Declaration of Helsinki and its later amendments.

Credit author statement

RT formal analysis of YTPA and MLAC, writing original draft, evolution of goals and aims, review and editing. EV formal analysis of CDSE, RS and MYSE, writing original draft, review and editing. MB secured funding, data acquisition, and review. TO conceptualization, methodology, review, secured funding, data acquisition, and supervision. All authors approved the final version.

References

Andreassen, T. W., Lorentzen, B. G., & Olsson, U. H. (2006). The impact of non-normality and estimation methods in SEM on satisfaction research in marketing. *Quality and Quantity*, 40(1), 39–58. https://doi.org/10.1007/s11135-005-4510-y

Bauman, A. A., Powell, B. J., Kohl, P. L., Tabak, R. G., Penalba, V., Proctor, E. E., Domenech-Rodriguez, M. M., & Cabassa, L. J. (2015). Cultural adaptation and implementation of evidence-Based parent-training: A systematic review and critique of guiding evidence. *Child and Youth Services Review*, 1(53), 113–120. https://.org/10.1016/j.childyouth.2015.03.025

Bergström, M., Cederblad, M., Håkansson, K., Jonsson, A. K., Munthe, C., Vinnerljung, B., Wirtberg, I., Östlund, P., & Sundell, K. (2020). Interventions in foster family care: A systematic review. *Research on Social Work Practice*, 30(1), 3–18. https://doi.org/10.1177/1049731519832101

Betz, N. E., Klein, K. L., & Taylor, K. M. (1996). Evaluation of a short form of the career decision-making self-efficacy scale. *Journal of Career Assessment*, 4(1), 47–57. https://doi.org/10.1177/106907279600400103

Blakeslee, J. E., Powers, L. E., Schmidt, J., Nelson, M., Fullerton, A., George, K., McHugh, E., Bryant, M., & Geenen, S., the Research Consortium to Increase the Success of Youth in Foster Care. (2020). Evaluating the my life self-determination model for older youth in foster care: Establishing efficacy and exploring moderation of response to intervention. *Children and Youth Services Review*, 119 105419. https://doi.org/10.1016/j.childyouth.2020.105419

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology*, 49(5), 713–734. https://doi.org/10.1177/0022022117749042

Bornman, J., Sevcik, R. A., Romski, M., & Pae, H. K. (2010). Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities*, 7(2), 111–118. https://doi.org/10.1111/j.1741-1130.2010.00254.x

Byrne, B. M. (2016). Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation*, 5(1), 51–65. https://doi.org/10.1037/ipp0000042

Castro, F. G., Barrera, M., Jr., & Martinez, C. R., Jr. (2006). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, 5(1), 1–5. https://doi.org/10.1023/B:PREV.0000013980.12412.cd

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005. https://doi.org/10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Ferrer-Wreder, L., Sundell, K., & Manssory, S. (2012). Tinkering with perfection: Theory development in the intervention cultural adaptation field. *Child Youth Care Forum*, 41(2), 149–171. https://doi.org/10.1007/s10566-011-9162-6

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Information Age Publishing.

Furukawa, E., Tangney, J., & Higashibara, F. (2012). Cross-cultural continuities and discontinuities in shame, guilt, and pride: A study of children residing in Japan, Korea and the USA. *Self and Identity*, 11(1), 90–113. https://doi.org/10.1080/15298868.2010.512748

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*, 1(1), 121–130. https://doi.org/10.1177/2515245917746500

Jovanovic, V., Cummins, R. A., Weinberg, M., Kaliterna, L., & Prizmic-Larsen, Z. (2019). Personal wellbeing index: A cross-cultural measurement invariance study across four countries. *Journal of Happiness Studies*, 20(3), 759–775. https://doi.org/10.1007/s10902-018-9966-2

Lai, K. (2019). More robust standard error and confidence interval for SEM parameters given incorrect model and nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(2), 260–279. https://doi.org/10.1080/10705511.2018.1505522

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance* (1st ed.). Routledge. https://doi.org/10.4324/9780203821961

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9(1), 93. https://doi.org/10.1037/1082-989X.9.1.93

Olsson, T. M., Blakeslee, J., Bergström, M., & Skoog, T. (2020). Exploring fit for the cultural adaptation of a self-determination model for youth transitioning from out-of-home care: A comparison of a sample of Swedish youth with two samples of American youth in out-of-home care. *Children and Youth Services Review*, 119. https://doi.org/10.1016/j.childyouth.2020.105484

Powell, D. A., & Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta-analysis. *Journal of Educational and Behavioral Statistics*, 26 (1), 105–132. https://doi.org/10.3102/10769986026001105

Powers, L. E., Turner, A., Westwood, D., Matiuszewski, J., Wilson, R., & Phillips, A. (2001). TAKE CHARGE for the future: A controlled field-test of a model to promote student involvement in transition planning. *Career Development and Transition for Exceptional Individuals*, 24(1), 89–104. https://doi.org/10.1177/088572880102400107

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria:R Foundation for Statistical Computing. https://www.R-project.org/

Rhemtulla, M., Brosseau-Laird, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Shi, D., Song, H., & Lewis, M. D. (2019). The impact of partial factorial invariance on cross-group comparisons. *Assessment*, 26(7), 1217–1233. https://doi.org/10.1177/1073191117711020

Sundell, K., Beelman, A., Hasson, H., & von Thiele Schwartz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child & Adolescent Psychology*, 45(6), 784–796. https://doi.org/10.1080/15374416.2015.1020540

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. https://doi.org/10.1177/109442810031002

von Thiele Schwarz, U., Aarons, G. A., & Hasson, H. (2019). The value equation: Three complementary propositions for reconciling fidelity and adaptation in evidence-based practice implementation. *BMC Health Services Research*, 19(1), 1–10. https://doi.org/10.1186/s12913-019-4668-y

Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement*, 1, 165–178. PMID: 7850498.

Wehmeyer, M. L., & Palmer, S. B. (2003). Adult outcomes for students with cognitive disabilities three years after high school: The impact of self-determination. *Education and Training in Developmental Disabilities*, 38(2), 131–144.

Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859