Westat

Improving lives through research™

Examining Construct Validity of a Child Health Interview Instrument

Authors

Matthew Finster
Jill Feldman

2019

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Abstract

Objectives: The objective of this study was to examine the construct validity and bias of a child health interview (CHI) instrument to assess young children's food and activity preferences.

Design: A quantitative study examining the construct validity and bias of a young child health interview instrument using confirmatory factor analysis (CFA) and multi-group CFA examining whether the measurement model of the instrument is equivalent (or invariant) across different child age groups.

Methods: We collected child interview data from 783 children (3-year-olds = 307; and 4-year-olds = 476) across 20 childcare sites in North Carolina during fall 2017 and spring 2018. To assess the factorial validity of the survey for children of different ages, we conducted group-specific CFAs using the 3- and 4-year-olds' responses and conducted a multi-group CFA.

Results: The model fit indices indicate the hypothesized structure of the model fits the data well and subscale factor loadings are significantly related to each of their respective domains. The multi-group analysis demonstrates configural and scalar invariance across the two groups (3- and 4-year-olds), indicating the two groups can be pooled together for further analysis.

Conclusion: This study presents some initial evidence of the construct validity of the CHI instrument. The results of the age-group-specific CFAs and multi-group CFA indicate there is no evidence of construct bias for young children of different ages.

Background

Obesity rates among children and adolescents in the United States have risen over the past 25 years (Ogden et al., 2016) and remain high despite attempts to counter them. Being overweight or obese in childhood is associated with negative school outcomes such as: reduced math and reading skills; decreased executive functioning skills; increased detention, absenteeism, and tardiness; and negative social-behavioral outcomes, including increased internalizing and externalizing behavior problems and ADHD (Datar & Sturm, 2006; Davis & Cooper, 2011; Pulgaron, 2013; Shore et al., 2008). Given the long-lasting and far-reaching negative effects associated with being overweight or obese in early childhood, childcare providers are uniquely positioned to cultivate habits leading to a healthy weight among the infants and toddlers in their care.

Links between children's physical health and academic achievement have been well-researched over the past decade. Numerous studies and meta-analyses demonstrate the importance of nutrition and fitness to academic achievement (Centers for Disease Control and Prevention, 2010; Fedewa & Ahn, 2011; Rampersaud, Pereira, Girard, Adams, & Metzl, 2005). Combined with an understanding of the relation between children's physical health and adult health outcomes, these findings have prompted calls for the integration of wellness initiatives within educational settings. To support the evaluation of wellness initiatives in early educational settings, this study examines the construct validity of a young child health interview instrument, to assess whether there is evidence of the factorial validity of the CHI and whether there is construct bias across groups of young children (3- and 4-year-olds).

Methods

Instrument

As part of a larger evaluation, a child health interview instrument (alternately referred to here as the instrument) was developed to assess the food and activity preferences of children enrolled in 20 childcare sites in four counties in North Carolina by staff at Westat (Feldman, Standing, Quintanilla, & Silva, 2017). The instrument was developed to provide a developmentally appropriate interview protocol for young children to provide baseline information and assess changes in young children's food and activity preferences over time. There are numerous validated nutrition and physical activity instruments for school-age children (e.g., Hoelscher, Day, Kelder, & Ward, 2003; Penkilo, George, & Hoelscher, 2008; Thiagarajah et al., 2006, 2008). Similar to other child health instruments (e.g., California Department of Public Health, 2017), the Child Health Interview (CHI) has nutrition and physical activity sections. The nutrition section is further broken out by fruits and vegetables. The interview includes questions paired with pictures of food items and young children from diverse backgrounds engaged in a range of developmentally appropriate activities. To ensure children recognized each activity and food before being asked if they liked them, assessors presented them with pictures of activities and foods and prompted the children to identify each one.

Children were asked how much they liked four fruits, four vegetables, water, and six activities at baseline and followup. The instrument scores children's preferences for fruits, vegetables, and physical activity for each item (0 = dislike, 1 = like). The scoring of the instrument is based on cumulative scores for the domains, which range from 0 to 4 for fruits and vegetables and 0 to 6 for physical activity. However, for the confirmatory factor analysis (CFA), we used the individual instrument item responses (i.e., dislike and like) to assess the factorial validity of the instrument, as opposed to the domain scores, which would not provide the factor loading for each interview item.

Participants and Methods

We collected child interview data from 783 children (3-year-olds = 307; and 4-year-olds= 476) across 20 childcare sites in four counties in North Carolina. We used Mplus Version 8 (Muthén &

Muthén, 2018) to conduct multiple group-specific (i.e., 3- and 4-year-olds) CFAs to examine the factor structure of the instrument and establish its construct validity or the extent to which the scales measure the intended underlying constructs in very young children. This empirical evidence indicates whether the instrument items are related to the underlying factors they are theorized to represent and whether relationships exist among items across factors. This evidence of factorial validity, combined with content validity, indicates an instrument's construct validity.

We examined the relationships among interview items and three subscales (food preferences – fruit; food preferences – vegetables; and activity preferences) using CFA to see whether findings support the existence of three distinct domains with items loading distinctly onto each one. The CFA tests the extent to which the hypothesized three-factorial model is supported empirically by data.

The CFAs use weighted least squares with means and variances adjusted (WLSMV) estimator, delta parameterization, and probit link to estimate the model fit indices and parameters. WLSMV is a robust estimator, which does not assume normally distributed data, thus providing the best option for modeling categorical or ordered data (Brown, 2006).

We also used multi-group CFA to examine the factorial equivalence (or invariance) of the instrument across groups of 3- and 4-year-olds. The purpose of the multi-group analysis is to test whether components of the measurement model are equivalent (also referred to as invariant) across the age groups (i.e., 3- and 4-year-olds). Testing for factorial equivalence across the groups can provide further evidence of construct validity by demonstrating that across two different groups, the survey items and their relationships to the domains are similar to each other. The general process of assessing measurement invariance often follows the testing of a series of increasingly restrictive hypotheses: as equivalence is demonstrated across each parameter, a higher level of invariance is met (Meredith, 1993; Steenkamp & Baumgartner, 1998). While there is an array of types of measurement invariance and a lack of agreed-upon terminology, Steenkamp and Baumgartner (1998) provide the following terms for levels of invariance: (1) configural invariance, (2) metric invariance, (3) scalar invariance, (4) factor covariance invariance, (5) factor variance invariance, and (6) error variance invariance. Since we are focused on construct validity in this analysis, not all of the invariance tests are of interest (Bryne, 2008), and with binary variables (using weighted least squares estimation and the delta parameterization), we can only test the configural and scalar models (Muthén & Muthén, 2015). The first level of measurement invariance is configural invariance, in which only the number of factors and loading patterns must be constant between groups, which is considered the least restrictive (or weakest) multi-group model (Horn & McArdle, 1992). The configural model serves two important functions: (1) it tests equivalence of the model parameters for the two groups simultaneously, and (2) it provides the baseline value against which other multi-group models are compared (Byrne, 2012). The second level of invariance is metric invariance, in which the equivalence of the factor loadings is tested. The third level of invariance is scalar invariance, in which, in addition to the parameters above, tests the equivalence of the measurement intercepts. Throughout the process, if the model fit is significantly worse when an additional parameter is held constant, that additional parameter is not equivalent (or invariant) across the two groups.

To test for factorial equivalence of the instrument across the two groups, we followed Byrne's approach (2012) by gradually restricting parameters of the measurement models. We followed the steps below to test for measurement invariance across the two groups: (1) establish group-specific baseline CFA models; (2) test the configural models (or baseline model); (3) test the scalar model by constraining factor loadings and intercepts (i.e., thresholds) so that they are equivalent; and (4) after demonstrating scalar equivalence across the two age groups, conduct a CFA using the pooled sample of 3- and 4-year-olds.

We started by conducting group-specific (3- and 4-year-olds) baseline CFA models of the instrument. We established the same group-specific baseline models for 3- and 4-year-olds and proceeded with testing the configural model. In this analysis, the configural model consists of three factors (i.e., fruits, vegetables, and physical activity), with fruits and vegetables each consisting of four items, and physical activity consisting of six items. After testing the configural model of the instrument using 3- and 4-year-olds' responses, we tested the scalar model (Muthén & Muthén, 2015). As mentioned above, typically, the next step after testing the configural model is to constrain the factor loadings across the two groups (referred to as the metric invariance or model); however, for models with binary variables (using weighted least squares estimation and the delta parameterization), only the configural and scalar model settings are available (Muthén & Muthén, 2015). In addition to restricting the number of factors and loading patterns to be equal, the scalar model tests the extent to which indicator intercepts (i.e., thresholds with binary data) are equivalent across the two groups (Muthén & Muthén, 2015). Based on the results of these tests, following Jöeskog's (1971) recommendation, we pooled the data of the 3- and 4-year-olds and conducted a CFA as a single group.

Results

Descriptive statistics for the number and proportion of children responding to items on the Child Health Interview (CHI) instrument are presented in Table 1. Across the three groups (3-year-olds, 4-year-olds, and both age groups combined—referred to as pooled), over 90 percent of children indicated they liked various fruits (i.e., items X1-X4) and physical activities (i.e., items X9-X14). However, across the three groups, fewer children reported liking various vegetables (i.e., indicators X5-X8), although a majority did, with the proportion of children reporting "yes" ranging from .54 (X7 for 4-year-olds) to .79 (X8 for 3-year-olds).

Table 1. **Proportions and Counts of Children's Responses on the Child Health Interview Instrument by Age**

Items	Response	3-year-olds (n = 307)		4-year-olds (n = 476)		Pooled (3- and 4-year-olds, n = 783)	
		Proportions	Counts	Proportions	Counts	Proportions	Counts
Likes bananas (X1)	yes	0.94	289	0.90	429	0.92	718
	no	0.06	18	0.10	47	0.08	65
Likes apples (X2)	yes	0.96	294	0.94	445	0.94	739
	no	0.04	13	0.07	31	0.06	44
Likes grapes (X3)	yes	0.96	295	0.95	454	0.96	749
	no	0.04	12	0.05	22	0.04	34
Likes oranges (X4)	yes	0.94	287	0.92	439	0.93	726
	no	0.07	20	0.08	37	0.07	57
Likes broccoli (X5)	yes	0.76	232	0.72	341	0.73	573
	no	0.24	75	0.28	135	0.27	210
Likes carrots (X6)	yes	0.79	241	0.76	361	0.77	602
	no	0.22	66	0.24	115	0.23	181
Likes tomatoes (X7)	yes	0.65	200	0.54	256	0.58	456
	no	0.35	107	0.46	220	0.42	327
Likes peas (X8)	yes	0.79	243	0.67	321	0.72	564
	no	0.21	64	0.33	155	0.28	219
Likes stretching (X9)	yes	0.87	267	0.84	400	0.85	667
	no	0.13	40	0.16	76	0.15	116
Likes jumping (X10)	yes	0.95	290	0.95	450	0.95	740
	no	0.06	17	0.06	26	0.06	43
Likes running (X11)	yes	0.95	292	0.94	449	0.95	741
	no	0.05	15	0.06	27	0.05	42
Likes swinging (X12)	yes	0.98	302	0.99	473	0.99	775
	no	0.02	5	0.01	3	0.01	8
Likes biking (X13)	yes	0.97	298	0.97	462	0.97	760
	no	0.03	9	0.03	14	0.03	23
Likes dancing (X14)	yes	0.94	289	0.91	434	0.92	723
	no	0.06	18	0.09	42	0.08	60

The CFA hypothesizes a three-factor model with fruits, vegetables, and physical activity each being a distinct factor. The goodness-of-fit indices for the group-specific CFAs indicate a reasonably good fit between the hypothesized model and observed data, based on model fit recommendations (e.g., Bentler, 1990; Browne & Cudeck, 1993; Hu & Bentler, 1999). (The fit indices for the pooled sample are presented in Table 2 but discussed below after the multigroup analysis demonstrates factorial equivalence across the two groups, indicating that the data should be pooled in a single-group CFA.) The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) for the group-specific 3-year-old model are within the acceptable range (between 0.90 to 0.95 [Bentler, 1990]) and well-fitting range (greater than .95 [Hu & Bentler, 1999]). The CFI and TLI for the group-specific 4-year-old model are both in the well-fitting range (greater than .95 [Hu & Bentler, 1999]). The root mean square error of approximation (RMSEA) absolute fit indices are also within the well-fitting range being below .05 (Browne & Cudeck, 1993). While the standardized root mean square residual (SRMR) is a commonly reported fit statistic (Brown, 2006), we do not report it since it is not recommended for use with binary outcomes (Yu, 2002). We also examined additional model parameters (e.g., allowing indicators to cross-load or covary) based on modification indices, but these did not result in improvement in model fit for both groups. Overall, the goodness-of-fit indices indicate the interview items do an adequate job of measuring the underlying constructs they were designed to measure.

Table 2. Fit Indices of Confirmatory Factor Analysis (CFA) Models of the Child Health Interview Instrument by Age

Fit Indices	CFA (3-year-olds, n=307)	CFA (4-year-olds, n = 476)	CFA (pooled, 3- and 4-year-olds, n = 783)
Number of free parameters	31	31	31
Chi-square test of model fit:			
Value	118.54	84.83	98.37
Degrees of freedom	74.00	74.00	74.00
p-value	<0.001	0.18	0.03
Chi-square test of model fit for the baseline model:			
Value	992.83	870.36	1796.62
Degrees of freedom	91.00	91.00	91.00
p-value	<0.001	<0.001	<0.001
Bentler comparative fit index (CFI)[1]	0.95	0.99	0.99
Tucker-Lewis index (TLI)[2]	0.94	0.98	0.98
Root mean square error of approximation (RMSEA):[3]			
Estimate	0.044	0.018	0.021
90 percent confidence interval	0.029—0.059	0.000—0.033	0.007—0.031
Probability RMSEA <=.05	0.728	1.00	1.00

[1]Values 0.90 to 0.95 are indicative of acceptable fit (Bentler, 1990) and values > .95 are indicative of a well-fit model (Hu & Bentler, 1999).
[2]Values 0.90 to 0.95 are indicative of acceptable fit (Bentler, 1990) and values > .95 are indicative of a well-fit model (Hu & Bentler, 1999).
[3]Values <0.08 are indicative of adequate fit (Browne & Cudeck, 1993) and values 0.80 to 0.10 are indicative of mediocre fit (MacCallum et al., 1996).
Notes. The estimator is weighted least squares means and variance adjusted (WLSMV).

All item loadings (parameter estimates) are statistically significant (p <.001) and a majority of items load onto the factors at values greater than 0.7 (12 items for the 3-year-olds group and 9 items for the 4-year-olds group), which is expected given the high percentage of positive responses to most items. According to Hair et al. (2010), factor loading values greater than 0.5 are acceptable and values equal to or greater than 0.7 are good. For 3-year-olds, the range of item loadings ranges for each factor was: (1) Fruit, 0.66 to 0.90, (2) Vegetables, 0.71 to 0.93, and (3) Physical activity, 0.54 to 0.81. For 4-year-olds, the range of item loadings for each factor was: (1) Fruit, 0.55 to 0.83, (2) Vegetables, 0.61 to 0.82, and (3) Physical activity, 0.44 to 0.93. Table 3 presents the standardized item loadings by each factor for each group and Table 4 presents the factor correlations of CFAs for each group (i.e., 3-year-olds, 4-year-olds, and pooled 3- and 4-year-olds).

After establishing the baseline CFA models for each group (i.e., 3-year-olds and 4-year-olds), we conducted multi-group analysis by testing the configural and scalar models. The results are presented in Table 5. The CFI, TLI, and RMSEA fit indices are within the well-fitting range (Byrne, 2012; Hu & Bentler, 1999). Since the difference in chi-square values for two nested models using the WLSMV chi-square values is not distributed as chi-square (Muthén & Muthén, 2015), to obtain the correct chi-square difference test, we followed a two-step process using derivatives from the analysis to compute the chi-square difference test (for more information, see Muthén & Muthén, 2015). The non-significant chi-square difference test (p = .49) indicates that the more restrictive model (scalar [H0] versus configural [H1] model) does not significantly worsen the model fit. Overall, the results indicate that the factor loadings and thresholds operate equivalently across the two groups (i.e., factor loadings and thresholds are invariant or equivalent).

Given the findings regarding equivalency, we proceeded by combining (pooling) the 3- and 4-year-olds' data and conducting a CFA on the pooled sample as a single group. The CFI, TLI, and RMSEA fit indices presented in Table 2 for the pooled sample are in the well-fitting range, with CFI and TLI values greater than .95 (Hu & Bentler, 1999) and an RMSEA value of below .05 (Browne & Cudeck, 1993). The CFA of the pooled sample is depicted in Figure 1.

Table 3. Standardized Factor Loading Estimates of Confirmatory Factor Analysis Models of the Child Health Interview by Age

Factors	Items	3-year-olds model (n=307)				4-year-olds model (n = 476)				3- and 4-year-olds pooled (n= 783)			
		Estimate	S.E.	Est./S.E.	P-Value	Estimate	S.E.	Est./S.E.	P-Value	Estimate	S.E.	Est./S.E.	P-Value
Fruits (F1) by	Likes bananas (X1)	0.76	0.09	8.78	<.001	0.72	0.08	8.59	<.001	0.73	0.06	11.40	<.001
	Likes apples (X2)	0.90	0.08	12.04	<.001	0.71	0.08	8.83	<.001	0.77	0.06	12.41	<.001
	Likes grapes (X3)	0.72	0.12	6.07	<.001	0.55	0.10	5.27	<.001	0.63	0.08	7.54	<.001
	Likes oranges (X4)	0.66	0.10	6.34	<.001	0.83	0.07	11.72	<.001	0.75	0.06	12.02	<.001
Vegetable (F2) by	Likes broccoli (X5)	0.83	0.05	16.55	<.001	0.78	0.05	16.12	<.001	0.81	0.04	22.73	<.001
	Likes carrots (X6)	0.72	0.06	11.51	<.001	0.82	0.05	18.19	<.001	0.78	0.04	20.97	<.001
	Likes tomatoes (X7)	0.75	0.06	12.57	<.001	0.61	0.06	10.00	<.001	0.68	0.04	15.29	<.001
	Likes peas (X8)	0.93	0.04	21.48	<.001	0.71	0.05	13.98	<.001	0.78	0.04	21.25	<.001
Physical activity (F3) by	Likes stretching (X9)	0.74	0.06	11.87	<.001	0.72	0.07	11.19	<.001	0.76	0.05	16.14	<.001
	Likes jumping (X10)	0.74	0.08	9.80	<.001	0.79	0.07	11.58	<.001	0.81	0.05	15.29	<.001
	Likes running (X11)	0.73	0.09	8.01	<.001	0.57	0.10	5.87	<.001	0.65	0.08	8.53	<.001
	Likes swinging (X12)	0.54	0.08	7.22	<.001	0.44	0.10	4.57	<.001	0.33	0.08	4.33	<.001
	Likes biking (X13)	0.81	0.09	8.71	<.001	0.49	0.13	3.91	<.001	0.58	0.10	5.69	<.001
	Likes dancing (X14)	0.72	0.08	8.59	<.001	0.93	0.06	16.52	<.001	0.87	0.05	18.57	<.001

Notes. The estimator is weighted least squares means and variance adjusted (WLSMV). Estimates presented are STDYX standardized, which is based on background and outcome variables. "By" is short for "measured by" and is used to indicate the regression estimate between the underlying factors (i.e., F1-F3) and the observed indicator variables (i.e., X1-X14).

Table 4. Factor Correlations of Confirmatory Factor Analysis Models of the Child Health Interview by Age

Factor	Factor	3-year-olds (n=307)				4-year-olds (n=476)				3- and 4-year-olds pooled (n = 783)			
		Estimate	S.E.	Est./S.E.	P-Value	Estimate	S.E.	Est./S.E.	P-Value	Estimate	S.E.	Est./S.E.	P-Value
Vegetable (F2) with	Fruit (F1)	0.82	0.07	11.66	<.001	0.65	0.08	8.45	<.001	0.71	0.06	12.58	<.001
Physical activity (F3) with	Fruit (F1)	0.86	0.10	8.38	<.001	0.49	0.10	5.11	<.001	0.61	0.08	7.97	<.001
	Vegetable (F2)	0.74	0.09	8.68	<.001	0.54	0.07	7.48	<.001	0.57	0.06	10.29	<.001

Notes. The estimator is weighted least squares means and variance adjusted (WLSMV). Estimates presented are STDYX standardized, which is based on background and outcome variables. “With” is short for “correlated with” and is used to indicate covariance relations between latent variables in the measurement model (i.e., F1-F3)

Table 5. Fit Statistics of Configural and Scalar Multigroup Models of the Child Health Interview Instrument

Tests of Model Fit	Configural	Scalar
Number of free parameters	62	54
Chi-square test of model fit		
Value	202.98	207.36
Degrees of freedom	148.00	156.00
p-value	.0018	.0037
Contribution from each group:		
Group 1 (Age 4, n = 476)	79.12	78.21
Group 2 (Age 3, n = 307)	123.86	129.16
Chi-square test for difference testing:		
Value	—	7.46
Δ Degrees of freedom	—	8
P-Value	—	0.49
Chi-square test of model fit for the baseline model:		
Value	1866.47	1866.47
Degrees of freedom	182	182
p-value	<.001	<.001
Bentler comparative fit index (CFI)[1]	0.97	0.97
Tucker-Lewis index (TLI)[2]	0.96	0.96
Root mean square error of approximation (RMSEA):[3]		
Estimate	0.03	0.03
90 percent C.I.	0.019—0.041	0.017 —0.039
Probability RMSEA <=.05	1	1

* p<.05.

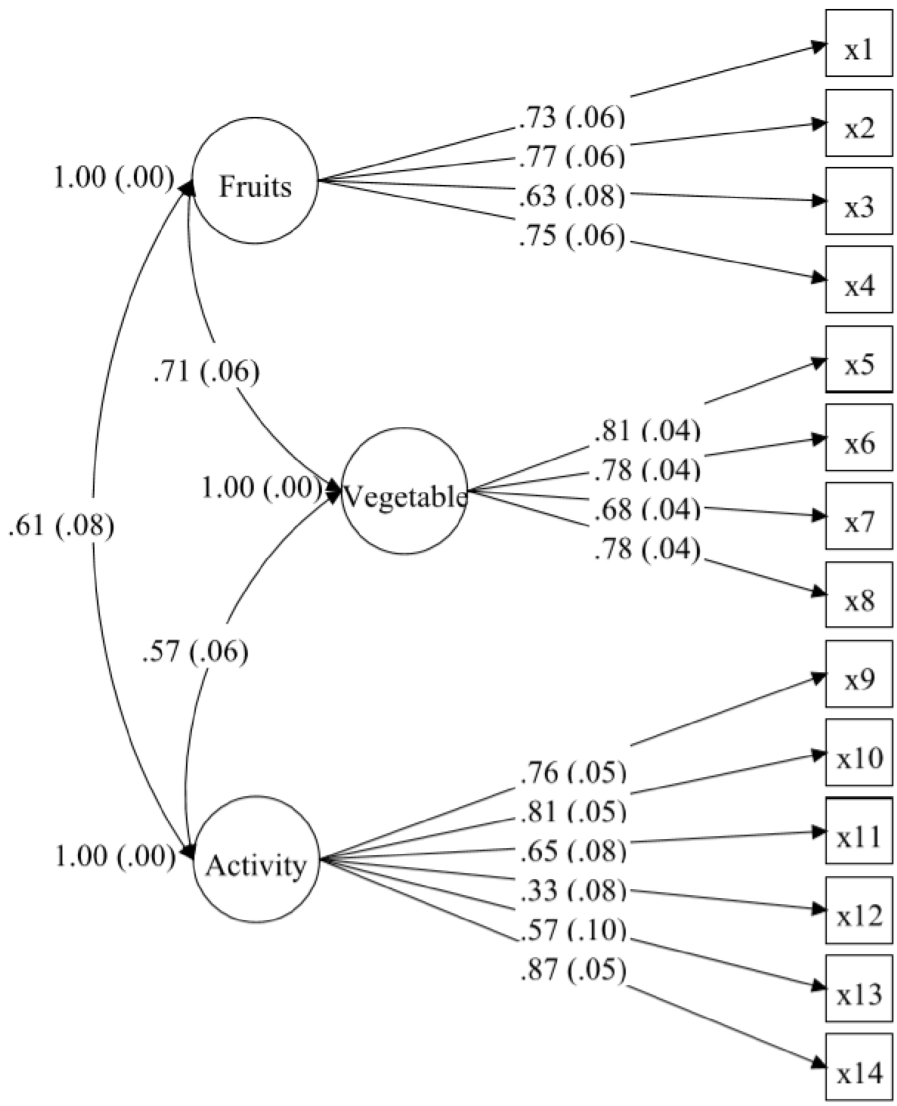
[1]Values 0.90 to 0.95 are indicative of acceptable fit (Bentler, 1990) and values > .95 are indicative of a well-fit model (Hu & Bentler, 1999).

[2] Values 0.90 to 0.95 are indicative of acceptable fit (Bentler, 1990) and values > .95 are indicative of a well-fit model (Hu & Bentler, 1999).

[3] Values <0.08 are indicative of adequate fit (Browne & Cudeck, 1993) and values 0.80 to 0.10 are indicative of mediocre fit (MacCallum et al., 1996).

Notes. N = 783. The estimator is weighted least squares means and variance adjusted (WLSMV). The chi-square value for WLSMV cannot be used for chi-square difference testing in the regular way. DIFFTEST option is used in Mplus to obtain a correct chi-square difference when WLSMV is used. Configural model has factor loadings and thresholds free across groups, residual variances fixed at one in all groups, and factor means fixed at zero in all groups. The scalar model has factor loadings and thresholds constrained to be equal across groups, residual variances fixed at one in one group and free in the other groups, and factor means fixed at zero in one group and free in the other groups.

Figure 1. **Confirmatory Factor Analysis Model of Child Health Interview Instrument Using the Pooled Sample**



Note. All factor loadings are significant at p<.001. Estimates are STDYX standardized, based on background and outcome variables. The single-headed arrows leading from each factor to the related items indicate the regression estimates of each item onto the underlying factor (i.e., the factor loadings). Standard errors are indicated in parentheses. The double-headed arrows from one factor to another indicate the factor correlations.

Discussion

This analysis presents initial evidence of the construct validity of the CHI. The analysis sought to determine the extent that the instrument items measured the factors of the instrument and assess the extent that the instrument operates equivalently across two different age groups (3- and 4-year-olds). The results of multi-group CFAs demonstrate that the factorial structure of the interview instruments is equivalent (or invariant) across 3- and 4-year-olds, indicating no construct bias across groups and it is appropriate to combine responses into a single-group CFA. The single-group CFA provides evidence of the factorial validity of the instrument and, combined with evidence of the content validity, demonstrates the construct validly of the child interview instrument. The results suggest it may provide a measure of young children's food and activity preferences and may be used as one measure of child-level health outcomes in evaluations of health programs.

Limitations

Establishing a measure's validity often requires the generation of ongoing evidence to support its use (Kane, 2006). This brief provides some evidence of the instrument's construct validity. Further evidence to substantiate the "validity" of the CHI should include data establishing the measure's convergent and predictive validity. Evidence of concurrent validity includes data that indicate scores on the interview are similar to scores on other instruments intended to measure young children's food and activity preferences. Predictive validity could be established by examining the extent to which interview responses are predictive of future scores on measures of child health like body mass index (BMI) scores, which indicate whether or not a child is at a healthy weight. In addition to examining other types of validity, it would also be beneficial to examine whether the instrument functions similarly across other subgroups of children (e.g., race and ethnicity), as evidence that the relationship between interview items and domains measured function similarly across groups, contributing further evidence of the reliability and validity of this interview instrument designed to provide young children with opportunities to express their food and activity preferences.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psichothema*, 20, 872–882.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* New York, NY: Routledge, Taylor & Francis Group.

California Department of Public Health. (2017). *Compendium of surveys: For nutrition education and obesity prevention.* Retrieved from https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/NEOPB/CDPH%20Document%20Library/RES_CompendiumofSurveys.pdf

Centers for Disease Control and Prevention. (2010). *The association between school-based physical activity, including physical education, and academic performance.* Atlanta, GA: U.S. Department of Health and Human Services.

Datar, A., & Sturm, R. (2006). Childhood overweight and elementary school outcomes. *International Journal of Obesity, 30*(9), 1449-1460.

Davis, C. L., & Cooper, S. (2011). Fitness, fatness, cognition, behavior, and academic achievement among overweight children: Do cross-sectional associations correspond to exercise trial outcomes? *Preventive Medicine, 52*, S65-S69.

Fedewa, A. L., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: A meta-analysis. *Research Quarterly for Exercise and Sport, 82*(3), 521-535.

Feldman, J., Standing, K., Quintanilla, P., & Silva, M. (2017). *Healthy starts for young children program: Child interview.* Rockville, MD: Westat.

Hair, J. F., Black, W. C., Balin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Harlow, England: Pearson Education Limited.

Hoelscher, D. M., Day, R. S., Kelder, S. H., & Ward, J. L. (2003). Reproducibility and validity of the secondary level School-Based Nutrition Monitoring student questionnaire. *Journal of the American Dietetic Association, 103,* 186-194.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement equivalence in aging research. *Experimental Aging Research, 18,* 117–144.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jöeskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: National Council on Measurement in Education and the American Council on Education.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement Issues and Practice, 14,* 5–8.

Muthén, B. O., & Muthén, L. K. (2018). Mplus (Version 8) [Computer software]. Los Angeles, CA: Author.

Muthén, B. O., & Muthén, L. K. (2015). VERSION 7.1 Mplus LANGUAGE ADDENDUM [Computer software]. Los Angeles, CA: Author.

Ogden, C. L., Carroll, M. D., Lawman, H. G., Fryar, C. D., Kruszon-Moran, D., Kit, B. K., & Flegal, K. M. (2016). Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *Journal of the American Medical Association, 315*(21), 2292-2299.

Penkilo, M., George, G. C., & Hoelscher, D. M. (2008). Reproducibility of the School-Based Nutrition Monitoring Questionnaire among fourth-grade students in Texas. *Journal of Nutrition Education and Behavior, 40* (American Cancer Society & 1996 Advisory Committee on Diet), 20-27.

Pulgaron, E. R. (2013). Childhood obesity: A review of increased risk for physical and psychological co-morbidities. *Clinical Therapy, 35*(1), 1-21.

Rampersaud, G. C., Pereira, M. A., Girard, B. L., Adams, J., & Metzl, J. D. (2005). Breakfast habits, nutritional status, body weight, and academic performance in children and adolescents. *Journal of the American Dietetic Association, 105*(5), 743-760.

Shore, S. M., Sachs, M. L., Lidicker, J. R., Brett, S. N., Wright, A. R., & Libonati, J. R. (2008). Decreased scholastic achievement in overweight middle school students. *Obesity, 16*(7), 1535-1538.

Steenkamp, J. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-107

Thiagarajah, K., Bai, Y., Lo, K., et al. (2006). P102: Assessing validity of food behavior questions from the School Physical Activity and Nutrition Questionnaire. *Journal of Nutrition Education and Behavior, 38*(4), S55-S56.

Thiagarajah, K., Fly, A. D., Hoelscher, D. M., et al. (2008). Validating the food behavior questions from the elementary school SPAN questionnaire. *Journal of Nutrition Education and Behavior, 40* (National Cancer Institute & 5 a Day Program Evaluation Group), 305-310.

Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation, University of California, Los Angeles). Retrieved from http://www.statmodel.com/download/Yudissertation.pdf