

Inter-regional metric disadvantages when comparing country happiness on a global scale. A Rasch-based consequential validity analysis

Desventajas métricas entre regiones al comparar la felicidad de los países a escala global. Un análisis Rasch de validez consecuencial

Diego Fernando Rojas-Gualdrón^{1*}

Abstract

Measurement confounding due to socioeconomic differences between world regions may bias the estimations of countries' happiness and global inequality. Potential implications of this bias have not been researched. In this study, the consequential validity of the Happy Planet Index, 2012 as an indicator of global inequality is evaluated from the Rasch measurement perspective. Differential Item Functioning by world region and bias in the estimated magnitude of inequalities were analyzed. The recalculated measure showed a good fit to Rasch model assumptions. The original index underestimated relative inequalities between world regions by 20%. DIF had no effect on relative measures but affected absolute measures by overestimating world average happiness and underestimating its variance. These findings suggest measurement confounding by unmeasured characteristics. Metric disadvantages must be adjusted to make fair comparisons. Public policy decisions based on biased estimations could have relevant negative consequences on people's health and well-being by not focusing efforts on real vulnerable populations.

Resumen

Confusión en la medición por diferencias socioeconómicas entre regiones del mundo podría sesgar las estimaciones de felicidad de los países y desigualdad global. Las implicaciones potenciales del sesgo no han sido investigadas. En este estudio, la validez consecuencial del Happy Planet Index, 2012 como indicador de desigualdad global es evaluada desde la perspectiva métrica de Rasch. Se analizaron Funcionamiento Diferencial de los Ítems por región del mundo y sesgo en la estimación de la magnitud de las desigualdades. El índice original subestimó la desigualdad relativa entre regiones del mundo en 20%. El FDI no tuvo efecto en las medidas relativas, pero afectó las absolutas sobreestimando el promedio mundial de felicidad y subestimando su varianza. Estos resultados sugieren confusión en la medición por características no observadas. Decisiones políticas basadas en estimaciones sesgadas podrían tener consecuencias negativas en la salud de las personas al no enfocar esfuerzos en las reales poblaciones vulnerables.

Keywords

Economics; global health; happiness; health inequalities; validation studies.

Palabras Clave

Desigualdades en la salud; economía; estudios de validación; felicidad; salud global.

¹ *Facultad de Medicina, Universidad CES, Medellín, Colombia.*

* **Corresponding author:** drojas3@unab.edu.co

Manuscript received 03-05-2017; revised 20-06-2017; accepted 20-07-2017.

1. Introduction

Country happiness is a current topic of research, public opinion and public policy. The use of happiness indicators has

been proposed along with economic growth indicators such as Gross Domestic Product to evaluate the welfare and progress of nations (Madalina, 2015; Musikanski, 2014). The common rationale of supporters is based on the limitations of economic

growth indicators as proxies of the welfare of a country and the need to define proper goals as happiness (Cummins, 2016). On the other hand, critics cite the conceptual and methodological issues in the definition and measurement of happiness at both the individual and the aggregate level (Stewart, 2014). Recent research in what has been termed “happiness economics” has raised awareness about the underlying relations of happiness with economic and non-economic policies and factors (Darma, 2013; Oxa, Arancibia, & Campero, 2014), and about the affordability of happiness policies, especially for low-income countries (Johns & Ormerod, 2007).

One dimension that has not been made explicit in this discussion relates to inequality. Questions about how happy a country should be or about how much difference between countries must be observed to consider it unjust have not been sufficiently researched. Even when some authors have considered similar questions (Gandelman & Porzecanski, 2013; W. Kalmijn & Veenhoven, 2014; W. M. Kalmijn & Arends, 2010), the scarcity of research on the validity of the measures of country happiness raises concerns about the conceptual and methodological support of empirical studies that estimate global inequalities in country happiness.

It is known that behavioral patterns influence dimensions of country happiness indicators such as life expectancy, ecological footprint and perceived well-being (Popova, 2014; Richards, Jiang, Chau, Bauman, & Ding, 2015; Stolarski, Jasielska, & Zajenkowski, 2015; Young-Chool & Ji-Hyun, 2016). However, the mechanisms of this influence and the prevalence of specific behavioral patterns are not constant between world regions (Cordero, Salinas-Jiménez, & Salinas-Jiménez, 2017). Moreover, geographical psychologists and human geographers have shown how people’s perceptions and behaviors are influenced by the places where they live (Han, 2015; Kim, Chun, & Sohn, 2015; Sujarwoto & Tampubolon, 2015). Consequently, differences between world regions in terms of cultural, social and economic characteristics may influence the performance of country happiness indicators within world regions, leading to measurement confounding in the estimation of both absolute measures of country happiness and the magnitude of global inequalities. Thus, the fundamental assumptions of exchangeability and collapsibility required for unbiased comparisons (Mansournia & Greenland, 2015) between world regions and between countries from different regions may not apply for measures of country happiness.

In psychometrics, measurement confounding has been studied from the perspective of Differential Item Functioning (DIF). DIF can be understood as the effect of a grouping characteristic (i.e. world region) on the relationship between the level of attribute and the response pattern in the items that constitute an index (Crane, Gibbons, Jolley, & van Belle, 2006). The presence of relevant DIF is an evidence of no collapsibility and no exchangeability as it indicates that the performance of a measure in the population is not invariant among subpopulations (Chapman, Fiscella, Duberstein, Kawachi, & Muennig, 2014). Relevant DIF may produce sub-

population specific metric (dis)advantages, introducing bias in the comparison of subpopulation measures. These subpopulation metric (dis)advantages can be considered themselves unfair and are informative of the consequential validity of a measure and its generalizability. Still, measures must be unbiased between world regions (i.e. relevant DIF-free) to obtain valid estimations of country happiness and to avoid wrong conclusions about the magnitude of global inequalities.

Reporting evidence of the consequential validity of country happiness measures may contribute to a better understanding of its usability as an alternative to classical economic indexes in global welfare and global inequality analyses. Thus, the objective of this study was to evaluate the consequential validity of the Happy Planet Index, 2012 (HPI), a country happiness index available for 151 countries (The New Economics Foundation, n. d.-a), as an indicator of global inequalities. For that purpose, inter-regional metric (dis)advantages were estimated by Differential Item Functioning and its impact on findings and conclusions was analyzed.

2. Methods

A transcultural Rasch-based consequential validity analysis of the Happy Planet Index was performed following an adaptation of Wolfe and Smith (2007) guidelines for measure validation studies.

2.1 Data source

Data on the Happy Planet Index was obtained from the project webpage (Foundation, n.d.). The dataset includes data from 151 countries across the world. The Thematic World Map was obtained from the Environmental Systems Research Institute ESRI community repository (ESRI, n.d.).

2.2 Instrument

2.2.1 Happy Planet Index

The HPI is a composite index developed by the New Economics Foundation which is supposed to measure country efficiency, or “how many long and happy lives each (country) produces per unit of environmental output”. The Index uses global data on Life expectancy, Experienced well-being and Ecological footprint and was calculated for the year 2012 (Foundation, 2012). According to the 2012 report the variables used for calculation are operationalized as follows:

Life expectancy: 2011 data taken from United Nations Development Programme - Human Development Report (Programme, 2011).

Experienced well-being: Arithmetic mean of individual responses to the Ladder of Life question in the Gallup World Poll (Gallup, n.d.). Latest data for each country as of February 2012.

Ecological footprint: 2008 data taken from Global Footprint Network for 142 countries (G. F. Network, n.d.). An original estimate was made for the nine remaining countries (Foundation, n.d.).

Additionally, the world region and sub-region of each country was considered following the classification of the original report: **1: Latin America: 1a:** Central America, Mexico and the Caribbean; **1b:** South America. **2: Western World: 2a:** Australia and New Zealand; **2b:** North America; **2c:** Western Europe; **2d:** Nordic Europe; **2e:** Southern Europe. **3: Middle East and North Africa: 3a:** North Africa; **3b:** Middle East and South West Asia. **4: Sub Saharan Africa; 4a:** Southern and Central Africa; **4b:** East Africa; **4c:** West Africa. **5a: South Asia. 6: East Asia: 6a:** China; **6b:** Wealthy East Asia; **6c:** South East Asia. **7: Transition States: 7a:** Central Asia and Caucasus; **7b:** Central and Eastern Europe; **7c:** Russia, Ukraine and Belarus.

2.3 Data analysis

The original data used to calculate the HPI was used to construct a Rasch measure of country happiness. Raw data was collapsed into ten categories using the formula (data-minimum)/range. Two Rasch measures were constructed, one based on raw data and a DIF-adjusted measure after analyzing their psychometric properties as follows:

2.3.1 Category function

Category function was assessed with Mean Square (MNSQ) and Measure->category/Category->Measure coherence statistics. Appropriate fit was considered for values between 0.5 – 2.0 for MNSQ and above 40% for coherence (Bond & Fox, 2015) Item measures β with their standard errors (expressed in logits) and the item-measure correlations are also reported. Local dependency was analyzed with standardized residual correlations, values > 0.7 were considered evidence of local dependency as they would indicate more than 50% of shared variance. Invariance of item measure between world regions was assessed by Differential Item Functioning techniques (Engelhard, 2013). For each country the DIF size against the global mean was calculated and tested with the two-sided t-test. Overall DIF was assessed for each item and tested with the item-measure chi-square test. Probabilities lower than 0.05 were considered statistically significant and DIF sizes greater than 0.5 logits were considered substantively significant. To adjust for inter-regional DIF without changing the raw scores, items were grouped into world region clusters according to DIF size in such a way that inter-regional overall DIF was not statistically significant.

2.3.2 Reliability and unidimensionality

Item reliability and separability are presented. Unidimensionality was assessed with principal component analysis of the residuals. Expected values were at least 50% of the variance explained by the Rasch measure and residual components < 2.0 eigenvalues (Bond & Fox, 2015). Country reliability and separability are also presented.

2.3.3 Consequential validity: estimation of inter-regional inequalities

The differences in the estimation of the inter-regional inequalities in happiness were compared by four techniques: 1) con-

trasting descriptive statistics obtained by the original HPI and the two Rasch measures (DIF-unadjusted and DIF-adjusted); 2) comparing the variance partition coefficients (VPC) estimated by random intercept linear regression variance decomposition analyses, where VPC expresses the percentage of variance attributable to inter-regional differences (Merlo, Wagner, Ghith, & Leckie, 2016); 3) graphically comparing the distribution of happiness according to HPI and DIF-adjusted Rasch measure of happiness on the world map; and 4) comparing the top 20 happiest countries according to HPI and the DIF-adjusted Rasch measure.

3. results

3.0.1 Category function

For the DIF-unadjusted Rasch measure the mean MNSQ were 0.91 ($SD = 0.23$) and 0.98 ($SD 0.24$) for Infit and Outfit respectively. Coherence was $< 40\%$ in five out of ten categories. After adjusting for inter-regional DIF, MNSQ means for Infit and Outfit changed to 0.98 ($SD 0.38$) and 1.03 ($SD 0.30$) and coherence was above 40% for all categories except the highest one which was estimated at 25%.

3.0.2 Item fit and differential item functioning

Item statistics and the overall between regions-DIF test are presented in Table 1. Both DIF-unadjusted and DIF-adjusted measures showed appropriate item fit according to Outfit MNSQ and item-measure correlations (> 0.70). For the DIF-unadjusted measure the three items showed an overall DIF according to the item-measure chi-square test. After creating new groups of world regions, as presented in Table 1, DIF was no longer statistically significant.

Compared to the world average difficulty of the item Footprint ($\beta = 2.06$), Sub-Saharan Africa and Transition states (regions 4 and 7) showed lower difficulty ($\beta = 1.37$). Latin America (region 1) ($\beta = 3.14$), and the other world regions ($\beta = 2.73$) showed higher difficulty than the world average. For Sub-Saharan Africa, Life expectancy difficulty ($\beta = -1.00$) was higher than for the other world regions ($\beta = -2.82$). For Latin America and Sub-Saharan Africa, Well-being difficulty ($\beta = -1.99$) was lower than world average ($\beta = -0.64$). This item is harder for the other regions ($\beta = -0.85$), particularly for the Transition states ($\beta = -0.59$).

Table 2 shows individual DIF size for each region compared with world average item difficulty. Sub-Saharan Africa showed absolute DIF size > 0.5 logits in the three categories, the Transition states showed absolute DIF size > 0.5 in Well-being and Footprint, and Latin America showed absolute DIF size > 0.5 only in footprint. Based on the DIF size, Footprint is underestimating Latin American countries and overestimating Sub-Saharan Africa and Transition state countries; Well-being is overestimating Latin America and Sub-Saharan Africa countries and underestimating Transition state countries; and Life expectancy is underestimating Sub-Saharan Africa countries.

Table 1. Item statistics and overall DIF for DIF-unadjusted and DIF-adjusted Rasch measures.

Item	DIF unadjusted Rasch measure				DIF-adjusted Rasch measure			
	β (ee)	Corr	Outfit	DIF- X^2 (prob.)	β (ee)	Outfit	DIF- X^2 (prob.)	
Footprint (1)					3.14 (0.26)	0.72	0.77	
Footprint	2.06 ^a (0.09)	0.83 ^a	1.01 ^a	25.50 ^a (.00)	2.73 ^b (0.15)	0.92 ^b	0.81 ^b	6.64 ^b (0.08)
Footprint (4, 7)					1.37 (0.17)	0.81	0.94	0.51 (0.47)
Well-being (7)					-0.59	0.69	0.89	
Well-being	- 0.64 ^a (0.08)	0.88 ^a	1.04 ^a	31.60 ^a (.00)	- 0.85 ^b (0.15)	0.89 ^b	1.40 ^b	3.61 ^b (0.30)
Life Expectancy Well-being (1, 4)					-1.00 -1.99 (0.14)	0.83 0.93	0.93 0.71	 0.46 (0.49)
Life Expectancy	- 1.43 ^a (0.08)	0.90 ^a	0.91 ^a	27.70 ^a (.00)	- 2.82 ^c (0.11)	0.79 ^c	1.53 ^c	 (0.91)

Note: **a**: All regions; **b**: Regions 2, 3, 5 and 6; **c**: Regions 1, 2, 3, 5, 6 and 7. **Latin America 1** :: **2**: Western World; **3**: Middle East and North Africa; **4**: Sub Saharan Africa; **5**: South Asia; **6**: East Asia; **7**: Transition States. **DIF**: Differential Item Functioning (in logits); β : Item difficulty (in logits); e.e. Standard error for β ; **Corr.**: Item-Trait correlation; **Outfit**: Outfit Mean-square; X^2 (**prob.**): Chi-square test of overall DIF.

3.0.3 Reliability and unidimensionality

DIF-unadjusted Rasch measure item reliability was 0.99 (17.90 strata). After adjusting for inter-regional DIF, item reliability was 0.99 (11.51 strata). In the case of countries, DIF-unadjusted reliability was around 0.85 – 0.90 (2.37 – 2.94 strata), and after adjusting for inter-regional DIF, country reliability increased to around 0.89 – 0.93 (2.84 – 3.51 strata). DIF-unadjusted and DIF-adjusted measures explained 87.3% and 90.5% of total variance. There was no evidence of other relevant measures according to the principal component analysis of residuals. In both cases the eigenvalues of contrast residuals were < 1.7 and the variance was explained by the first contrast of 6.7% and 2.2% percent respectively.

3.0.4 Consequential validity: estimation of inter-regional inequalities

Table 3 shows the descriptive statistics and variance decomposition results for original HPI and for both DIF-unadjusted and DIF-adjusted Rasch measures. HPI and Rasch measures are not directly comparable as they are expressed in different scales. However, HPI results show Latin America, South Asia and East Asia regions as the three happiest regions with small differences between them, while both Rasch measures have the Western World as the happiest region with a distance greater than two logits from the second region, which in both cases is Latin America. According to descriptive statistics (Table 3), DIF by world region overestimated (DIF-unadjusted > DIF-adjusted) Sub-Saharan Africa, South Asia, the Transition states and Latin America, and underestimated (DIF-unadjusted < DIF-adjusted) the Western world coun-

tries. Global average country happiness is also overestimated by the unadjusted-DIF measure.

Between-country variance (total variance in Table 3) is underestimated by DIF-unadjusted measure. However, when estimating the relative variability due to inter-regional differences in country happiness through the VPC (Table 3), both Rasch measures (DIF-unadjusted and DIF-adjusted) calculate that 63% of country differences in happiness are explained by inter-regional variability. Original HPI variance is not comparable with Rasch variances. However, relative measures like VPC are comparable. The original HPI calculates that 44% of country differences are explained by inter-regional variability. The original HPI underestimate the magnitude of inequalities between world regions.

Figure 1 shows sub-region averages of country happiness for both DIF-unadjusted (figure 1.a) and DIF-adjusted (figure 1.b) Rasch measures. The relative global distribution of happiness is similar according to both measures (i.e. similar map pattern). However, sub-region average happiness differed after adjusting for inter-regional DIF. Only China (sub-region 6a) and all the sub-regions of the Western World (sub-regions 2a–2e) got higher values (between 0.01 and 0.73 logits higher) after adjusting for inter-regional DIF (figure 1.b); all other sub-regions got lower values (between 0.30 and 1.31 logits higher).

When comparing measures at the level of individual countries (Table 4), the Happy Planet Index and DIF-adjusted Rasch measure differ significantly. According to HPI, 13 of the top 20 happiest countries are from Latin America (region 1), five from South and East Asia (regions 5 and 6), one

Table 2. Differential Item Functioning size by region compared to world average.

Id	World region	Life Expectancy		Well-being		Footprint	
		DIF	Prob.	DIF	Prob.	DIF	Prob.
1	Latin America	-0.04	0.84	-0.43	0.03	0.72	0.01
2	Western World	-0.52	0.10	0.04	0.86	0.16	0.37
3	Middle East and North Africa	-0.26	0.25	0.43	0.06	-0.23	0.35
4	Sub Saharan Africa	0.89	0.00	-0.65	0.00	-0.60	0.02
5	South Asia	-0.34	0.36	-0.17	0.68	1.81	0.09
6	East Asia	-0.55	0.09	0.29	0.30	0.29	0.40
7	Transition States	-0.22	0.21	0.57	0.00	-0.57	0.01

Note: **DIF:** Differential Item Functioning Size (in logits); **Prob:** two-sided t test probability.

Table 3. Descriptive statistics and variance decomposition of regional happiness according to the Happy Planet Index and Rasch estimates.

World region	Happy Planet Index		Rasch measure			
	Mean	SD	DIF-unadjusted		DIF-adjusted	
			Mean	SD	Mean	SD
Latin America	52.58	7.50	0.19	0.83	-0.42	1.24
Western World	43.35	5.14	2.07	0.86	2.55	1.21
Middle East/North Africa	42.59	8.75	-0.06	1.71	-0.47	2.41
Sub Saharan Africa	33.45	5.14	-2.79	1.02	-4.09	1.18
South Asia	50.08	4.29	-1.44	0.36	-2.46	0.49
East Asia	47.08	6.91	-0.03	1.18	-0.44	1.69
Transition States	40.02	6.09	-0.29	0.67	-1.09	0.91
World	42.24	9.08	-0.45	1.89	-1.06	2.55
Intercept		44.01		-0.33		-0.91
Inter-regional variance		33.92		1.90		3.53
Intra-regional variance		43.13		1.12		2.07
Total variance		77.04		3.02		5.60
Variance Partition Coefficient		44.02%		62.91%		63.03%

Note: **DIF:** Differential Item Functioning; **SD:** Standard deviation. Rasch measures in logits.

from Middle East/South West Asia (sub-region 3b), and one from the Transition States (region 7). According to the DIF-adjusted Rasch measure, 16 of the top 20 countries are from the Western world (region 2) and the remaining four are from Middle East/South West Asia (sub-region 3b). Only Israel is included in both top 20 lists.

4. Discussion

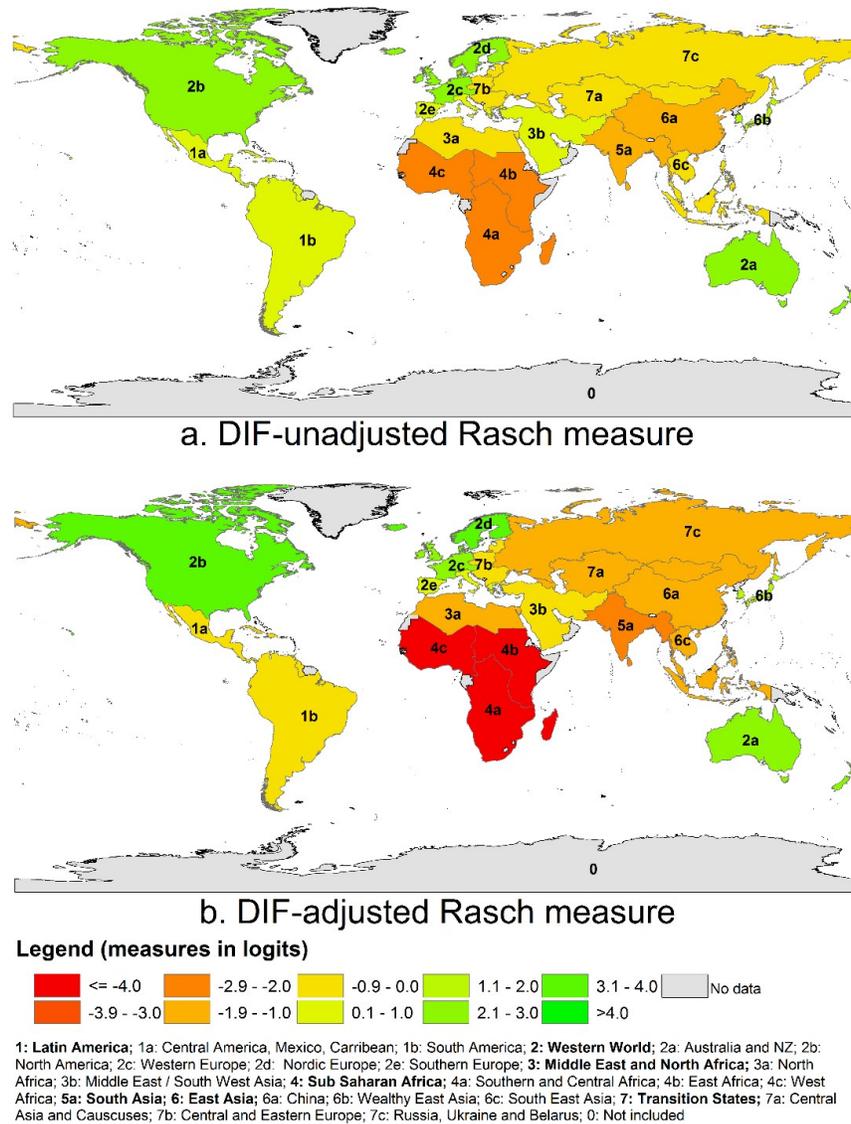
The objective of this study was to evaluate the consequential validity of the Happy Planet Index, 2012 (Foundation, n.d.) as an indicator of global inequalities, with special focus on inter-regional metric (dis)advantages.

HPI underestimated the relative magnitude of global inequalities in country happiness as explained by differences between world regions by about 20% compared to Rasch measures. HPI also made a disordered ranking of world regions and underestimated the absolute differences between them (and consequently between countries). HPI ranked Latin America and South Asia as the happiest world regions with a difference of 5.77 points, equivalent to 0.66 standard deviations.

The DIF-adjusted Rasch measure ranked the Western world and Latin America as the happiest world regions with a difference of 2.97 logits, equivalent to 1.25 standard deviations.

These differences between HPI and the DIF-adjusted measure suggest that HPI is influenced by measurement confounding by world regions. This bias could lead to major consequential validity implications. If country happiness measures are going to be considered alongside Gross Domestic Product or the Human Development Index as relevant indicators of welfare and progress, then sufficient validity arguments of their ability to perform unbiased inter-regional comparisons are mandatory. The underestimation of the magnitude of between-country inequalities explained by differences between world regions (i.e. VPC) could be wrongly interpreted as showing a decreasing gap that represents real achievements in global equity agendas. In a similar fashion, and considering that the best observed result is frequently used as the goal criteria in public policy analyses (Kanie & Biermann, 2017), misestimation of absolute differences between world regions

Figure 1. Comparison of sub-region average happiness estimated by DIF-unadjusted and DIF-adjusted Rasch measures.



(and countries) could erroneously lead to the setting of goals that are either too hard or too easy to achieve.

For example, according to HPI, Western countries could set a goal of improving their happiness to the levels observed in Latin America. However, inter-regional unbiased Rasch measures showed that the Western World's happiness is around 3 logits higher than in Latin America. Additionally, Rasch-measure based rankings matched with the rankings produced for the 2016 World happiness report (T. S. D. S. Network, 2016) for 16 out of 20 happiest countries. The original HPI only had three matches with the 2016 report. Moreover, a study conducted by (Jablonsky, 2013), using data envelopment analysis models, found similar results related to the disordered ranking of countries made by HPI.

DIF findings can only be compared between DIF-unadjusted and DIF-adjusted measures. Both measures showed a proper fit to Rasch model assumptions according to category function,

item and country fit statistics, reliability and unidimensionality analyses. However, Inter-regional metric (dis)advantages were identified through Differential Item Functioning. These differences did not affect the estimation of the relative magnitude of inter-regional inequalities (i.e. the VPC and rankings). However, the DIF-unadjusted measure overestimated world average happiness and underestimated its variance. These findings allow the hypothesis that Differential Item Functioning by world regions affects absolute but not relative estimations of inter-regional inequality in country happiness, meaning that the effect of DIF was proportional to DIF-unadjusted measures (Kjellsson & Petrie, 2017). Further studies to differentiate uniform and non-uniform DIF patterns by world region are required.

The fact that Latin America, Sub-Saharan Africa and the Transition States were the only world regions whose item difficulties significantly differed from the world average is

Table 4. World Ranking of the 20 happiest countries according to the Happy Planet Index and DIF-adjusted Rasch measure.

Happy Planet Index		DIF-adjusted Rasch measure			
Rank	Sub-region	Country	Rank	Sub-region	Country
1	1a	Costa Rica	1	2c	Luxembourg
2	6c	Vietnam	1	3b	Qatar
3	1b	Colombia	3	2d	Denmark
4	1a	Belize	4	2a	Australia
5	1a	El Salvador	4	2b	Canada
6	1a	Jamaica	4	2c	Netherlands
7	1a	Panama	4	2c	Switzerland
8	1a	Nicaragua	4	3b	United Arab Emirates
9	1b	Venezuela	9	2d	Finland
10	1a	Guatemala	9	2d	Iceland
11	5a	Bangladesh	9	2c	Ireland
12	1a	Cuba	9	3b	Kuwait
13	1a	Honduras	9	2d	Sweden
14	6c	Indonesia	9	2b	United States of America
15	3b	Israel	15	2c	Austria
16	5a	Pakistan	15	2c	Belgium
17	1b	Argentina	15	2c	France
18	7b	Albania	15	3b	Israel
19	1b	Chile	15	2d	Norway
20	6c	Thailand	20	2a	New Zealand

Note: **DIF**: Differential Item Functioning.

also worth mentioning.

Sub-Saharan countries have the lowest life expectancy worldwide, and have made the slowest progress in improving it (Caporale & Gil-Alana, 2016). For 2015, life expectancy in Africa was 60 years compared to 71.4 years worldwide. Some Sub-Saharan countries like Nigeria and Angola have a life expectancy of less than 55 years (Organization, n.d.). This disproportionate difference explains the 1.0 logit DIF size in life expectancy item. However, lower difficulties of around -0.6 logits in Well-being and Footprint weighted DIF and explain why Sub-Saharan African countries were overestimated by the DIF-unadjusted measure.

The higher difficulty in Well-being in transition states may be explained by living conditions that characterize nations transitioning to democratic governments. However, as this item is the only one of HPI based on subjective perceptions, an alternative explanation may be related to the way people define and perceive well-being in post-communist countries (Kelley & Evans, 2017). Also, it is relevant to consider that those countries were labeled as transition states because they started a transition process—not because of this process' result (Evans, 2011).

For Latin America, the Ecological Footprint difficulty is higher than for the worldwide measure. This means that Latin American countries must make greater progress than other countries to improve their happiness measure. Reasons for this DIF are not clear. Because ecological footprint is a composed index, global hectares per capita and their relation to country

happiness has not been studied as much as Life expectancy or Well-being. However, Footprint is the hardest item for all world regions and the only one harder than average (positive value).

The present study had some limitations. Even when comparisons of relative estimations allowed us to evaluate the consequential validity of HPI to measure inter-regional inequalities, the main limitation of this study was the impossibility of directly comparing absolute estimations made by original HPI and Rasch measures because they have different measurement units. Additionally, HPI was estimated directly from raw formats of items while Rasch measures were estimated after collapsing raw item scores to percentile based categories. The imprecision of information implied by categorization reduces the estimations of variance in Rasch measures (Engelhard, 2013). Nonetheless, reliability and model fit were good for a three-item measure tested in a sample of 151 countries. Item and country sample may seem small, however the low availability of quality data worldwide, especially in Low Income countries, must be considered when judging sample size.

This study also had some strengths. As far as the author could verify, this is the first validity analysis of country happiness measures that considers several validity facets and one of the few Rasch analyses explicitly oriented to consequential validity with a health inequality analysis. Psychometrics, and especially Rasch models have been widely used in health measurement. Nonetheless, their use in population health

metrics is an emerging research field. Another strength of this study was the use of several inequality analysis techniques to compare results and conclusions between the HPI and Rasch measures. This leads to robust evidence about the generalizability and consequential validity of country happiness comparisons, recognizing the differences between absolute and relative analyses of inequality.

A new version of HPI was released with data from 140 countries for 2016 (Foundation, n.d.). As in 2012 estimations, in this new version the 10 happiest countries come from Latin America and Asia Pacific. Considering the limitations identified in HPI, 2012 for the interpretation of HPI, 2016 is highly recommended. The additional analysis of potential DIF by official classification of countries such as World Bank income levels or world regions is required to a deepen our understanding of measurement confounding by unmeasured characteristics that differ between world regions.

5. Conclusion

Differential Item Functioning by world regions had a strong effect on the relative and absolute magnitude of inequalities, suggesting measurement confounding by unmeasured characteristics. The potential implications of this bias in the actions and decisions based on HPI must be carefully considered. Biased estimations could have relevant negative consequences on people's health and well-being by not focusing efforts on the truly vulnerable populations. More research and applied studies on the application of psychometric techniques in population health measurement, especially in health inequality assessment, is encouraged to improve evidence on the validity of inferences in this area.

References

- Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences (third edition)*. New York, NY: Taylor and Francis Group.
- Caporale, G. M., & Gil-Alana, L. A. (2016). Estimating time trends in life expectancy in sub-saharan africa. *Population Review*, 55(2), 1-24.
- Chapman, B., Fiscella, K., Duberstein, P., Kawachi, I., & Muennig, P. (2014). Measurement confounding affects the extent to which verbal iq explains social gradients in mortality. *Journal of Epidemiology and Community Health*, 68(8), 728-733.
- Cordero, J. M., Salinas-Jiménez, J., & Salinas-Jiménez, M. M. (2017). Interfaces with other disciplines: Exploring factors affecting the level of happiness across countries: A conditional robust nonparametric frontier analysis. *European Journal of Operational Research*, 256, 663-672.
- Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques. *difdetect* and *difwithpar*. *Medical Care*, 44(11 Suppl 3), S115-S123.
- Cummins, R. (2016). Happiness is the right metric to measure good societal functioning. *Society*, 53(3), 273-277.
- Darma, M. (2013). On the economics of happiness: the influence of income and non-income factors on happiness. *South African Journal of Economic and Management Sciences*, 16(1), 39-51. doi: 10.4102/sajems.v16i1.204
- Engelhard, G. (2013). *Invariant measurement: using rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge.
- ESRI, E. S. R. I. (n.d.). *Environmental systems research institute* [Web Page]. Retrieved from <http://www.esri.com/>
- Evans, A. B. (2011). The failure of democratization in russia: A comparative perspective. *Journal of Eurasian Studies*, 2(1), 40-51.
- Foundation, T. N. E. (n.d.). *Happy planet index*. Retrieved from <http://www.happyplanetindex.org/data/>
- Foundation, T. N. E. (2012). *The happy planet index: 2012 report*. Retrieved from <http://www.happyplanetindex.org/assets/happy-planet-index-report.pdf>
- Gallup, G. W. P. W. (n.d.). *Gallup world poll website* [Web Page]. Retrieved from <http://www.gallup.com/services/170945/world-poll.aspx>
- Gandelman, N., & Porzecanski, R. (2013). Happiness inequality: How much is reasonable? *Social Indicators Research*, 110(1), 257-269.
- Han, S. (2015). Social capital and subjective happiness: Which contexts matter? *Journal of Happiness Studies*, 16(1), 241-255.
- Jablonsky, J. (2013). Re-calculation of happy planet index using dea models. *Multiple Criteria Decision Making*(8), 56-66.
- Johns, H., & Ormerod, P. (2007). *Happiness, economics and public policy*. Westminster: IEA. Retrieved from <http://www.iea.org.uk/sites/default/files/publications/files/upldbook416pdf.pdf>
- Kalmijn, W., & Veenhoven, R. (2014). Index of inequality-adjusted happiness (iah) improved: A research note. *Journal of Happiness Studies*, 15(6), 1259-1265.
- Kalmijn, W. M., & Arends, L. R. (2010). Measures of inequality: Application to happiness in nations. *Social Indicators Research*, 99(1), 147-162.
- Kanie, N., & Biermann, F. (2017). *Governing through goals: Sustainable development goals as governance innovation*. Cambridge, Massachusetts: The MIT Press.
- Kelley, J., & Evans, M. (2017). The new income inequality and well-being paradigm: Inequality has no effect on happiness in rich nations and normal times, varied effects in extraordinary circumstances, increases happiness in poor nations, and interacts with individuals' perceptions, attitudes, politics, and expectations for the

- future. *Social Science Research*, 62, 39 - 74. doi: <http://dx.doi.org/10.1016/j.ssresearch.2016.12.007>
- Kim, K. H., Chun, J.-H., & Sohn, H. S. (2015). The influence of regional deprivation index on personal happiness using multilevel analysis. *Epidemiology and Health*, 37(e2015019). doi: 10.4178/epih/e2015019
- Kjellsson, G., & Petrie, D. (2017). Graphical tools for monitoring inequality: The beauty can lie (in the details). *Epidemiology*, 28(4), 604-607. doi: 10.1097/EDE.0000000000000672
- Madalina, C. (2015). Gdp vs. gnh. a new paradigm. *Annals of "Constantin Brancusi" University of Targu-Jiu Economy Series*, 349-354.
- Mansournia, M. A., & Greenland, S. (2015). The relation of collapsibility and confounding to faithfulness and stability. *Epidemiology*, 26(4), 466-472.
- Merlo, J., Wagner, P., Ghith, N., & Leckie, G. (2016). An original stepwise multilevel logistic regression analysis of discriminatory accuracy: The case of neighborhoods and health. *PLoS One*, 11(4), e0153778.
- Musikanski, L. (2014). Happiness in public policy. *Journal of Social Change*, 6(1), 55-85. doi: 10.5590/JOSC.2014.06.1.06
- Network, G. F. (n.d.). *Global footprint network website* [Web Page]. Retrieved from <http://www.footprintnetwork.org/es/>
- Network, T. S. D. S. (2016). *World happiness report, 2016 volume i* [Web Page]. Retrieved from http://worldhappiness.report/wp-content/uploads/sites/2/2016/03/HR-V1_web.pdf
- Organization, W. H. (n.d.). *Global health observatory (gho) data*. Retrieved from http://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends/en/
- Oxa, A., Arancibia, C., & Campero, S. (2014). Economía de la felicidad: evidencia empírica para latinoamérica. *Revista Perspectivas*, 34, 159-180.
- Popova, O. (2014). Can religion insure against aggregate shocks to happiness? the case of transition countries. *Journal of Comparative Economics*, 42, 804-818.
- Programme, U. N. D. (2011). *Sustainability and equity: a better future for all*. Washington DC. Retrieved from http://hdr.undp.org/sites/default/files/reports/271/hdr_2011_en_complete.pdf
- Richards, J., Jiang, K. X., Chau, P., Bauman, J., & Ding, A. D. (2015). Don't worry, be happy: cross-sectional associations between physical activity and happiness in 15 european countries. *BMC Public Health*, 15(1), 1-8.
- Stewart, F. (2014). Against happiness: A critical appraisal of the use of measures of happiness for evaluating progress in development. *Journal of Human Development & Capabilities*, 15(4), 293-307.
- Stolarski, M., Jasielska, D., & Zajenkowski, M. (2015). Are all smart nations happier? country aggregate iq predicts happiness, but the relationship is moderated by individualism-collectivism. *Intelligence*, 50, 153-158.
- Sujarwoto, S., & Tampubolon, G. (2015). Decentralisation and citizen happiness: A multilevel analysis of self-rated happiness in indonesia. *Journal of Happiness Studies*, 16(2), 455-475.
- Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using rasch models: part ii-validation activities. *Journal of Applied Measurement*, 8(2), 204-234.
- Young-Chool, C., & Ji-Hyun, J. (2016). Conditions associated with adolescents' happiness in oecd countries. *Journal of Economics & Economic Education Research*, 17(2), 17-28.