

RESEARCH ARTICLE

Testing the generalized validity of the Emotion Knowledge test scores

Ana R. Delgado^{1*}, Debora I. Burin², Gerardo Prieto¹

1 Facultad de Psicología, Universidad de Salamanca, Salamanca, Spain, **2** Facultad de Psicología, Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina

* adelgado@usal.es



Abstract

Differential item functioning (DIF) is of the utmost importance in order to corroborate the generalized validity of test scores in different groups. DIF indicates that an item does not function equally in different groups such as age, gender or cultural ones. Our objective was to contrast the generalized validity of the Emotion Knowledge (EK) test scores in a heterogeneous Argentinian sample composed of 100 females and 100 males (age range: 18–65). Data from the original validation sample (200 Spanish participants, half of them males) were conjointly analyzed (total $n = 400$). Results of the Rasch Model (RM) analysis indicated that both fit to the RM and reliability ($ISR = .97$, $PSR = .80$) were adequate. Item *logit* measures ranged from -3.89 to 3.68, and person *logit* measures ranged from -1.12 to 5.09, with a mean value of 2.36. DIF was tested for gender, age, educational level and country, with a few item contrasts found to be statistically significant. Even though small significant differences in EK scores were associated with educational level ($d = .25$) and country ($d = -.25$), they became non-significant after removing the seven country-related DIF affected items. We can conclude that there is enough evidence for the generalized validity of EK test scores in Argentina. Given that recent theories of human emotion consider conceptual knowledge supported by language as *constitutive* of emotions, the EK test can be used in academic or applied settings where individual differences in emotional competence might be relevant.

OPEN ACCESS

Citation: Delgado AR, Burin DI, Prieto G (2018) Testing the generalized validity of the Emotion Knowledge test scores. PLoS ONE 13(11): e0207335. <https://doi.org/10.1371/journal.pone.0207335>

Editor: Karl Bang Christensen, University of Copenhagen, DENMARK

Received: September 21, 2017

Accepted: October 30, 2018

Published: November 14, 2018

Copyright: © 2018 Delgado et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under Grant PSI2014-52369-P to Ana R. Delgado. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Recent theories of human emotion consider conceptual knowledge supported by language as *constitutive* of emotions [1–5]. In this view, emotions are not modules in the brain that trigger fixed expressive responses [6], but constructed affective states, guided by categories and language. Previous constructionist approaches conceived of emotions as semantic scripts of prototypical behaviors, expressions, labels and words [7]. Developmentally, children would go from a broad, valence-based system to knowing full scripts for specific discrete categories of emotion [8]. Furthermore, for the conceptual act theory [4] emotional categories are not fixed scripts, but constructed mental phenomena anchored in concepts and language. Emotions, like the rest of mental life, emerge as a consequence of the human brain's tendency to categorize, to make the contingencies meaningful. Different instances of sensory inputs, core affective states (valence, arousal), interactions, and behavior could be grouped together into the same category and given the same

name. Some of these categories might be cross-culturally stable, whereas other categories are culture specific. Language plays a central role in this view: words are the “glue” that brings together different instances into a coherent category [2, 4]. Therefore, the conceptual act theory predicts general agreement within broad emotional categories for people using the same language, even though certain sub-cultural differences in Emotion Knowledge (EK) could be found. In the context of discrete emotion theories, EK has been defined as related to the *understanding* of discrete emotions and differentiated from semantically close concepts such as emotion utilization (the adaptive use of emotion arousal) and emotion regulation [9].

The construction and validation of EK tests is of interest both from the theoretical and the applied points of view. The most-used emotional intelligence test is the Mayer Salovey Caruso Emotional Intelligence Test (MSCEIT), although only one of its facets, that of *understanding*, has received enough empirical support as a measure of aptitude [10]. Mayer, Salovey and Caruso [11] have clarified their original description of the understanding area of the MSCEIT: “[...] we meant that a person who possessed emotional knowledge could understand emotional word meanings and concepts, understand the situations [...]” (p. 404). They have recently described Emotional Intelligence as one of the broad intelligences in the context of a hierarchical model that empirically categorizes human abilities into areas such as fluid reasoning, visual spatial processing or *comprehension-knowledge*, considering that if emotional intelligence is really a discrete intelligence, it would be needed to make the case that there has evolved a separate reasoning capacity to *understand* emotions [12]. In addition to the relevance that Mayer, Salovey and Caruso attribute to EK [11, 12], emotional competence test scores predict various socially relevant outcomes [13–15].

The reasons summarized above led to the construction of language-based EK tests [16] by means of the Rasch Model (RM), an implementation of the invariant measurement approach [17–22]. The RM indicates that the probability that person n passes item i is $P_{ni} = \exp(B_n - D_i) / (1 + \exp[B_n - D_i])$, B_n : person level, D_i : item location. If the empirical data fit the model adequately, then person measures and item locations can be jointly measured on an *interval* scale in *logit* units. Evidence of unidimensionality was found when scaling the scores from the three EK tests conjointly [16] and so, for the purposes of this paper, we will refer to the EK *test*. In the invariant measurement realm, an important empirical testing of generalized validity can be carried out by testing the lack of Differential Item Functioning (DIF).

DIF indicates that an item measures differently in different contexts: item locations are not invariant across various groups, breaking the model requirement of person invariant calibration of test items [22]. It is unlikely to be detected at an individual level, and so it is usually checked for groups based on gender or culture to ensure test fairness [23]. DIF analysis tests the generalized validity of the measures for different groups. The usual procedure in the RM context is to test the standardized difference between item calibrations in two groups (i.e., Argentina and Spain, male and female, etc.) with Bonferroni-corrected *alpha* levels; the Rasch-modeled scores from the analysis of all the participants are held constant, providing the joint measurement scale in *logit* units [24, 25].

Thus, the objective of this study was to test the generalized validity of the Emotion Knowledge (EK) test scores, originally validated in Spain, with new data from an Argentinian sample. Our aim was achieved.

Materials and methods

Participants

The sample was composed of 100 females and 100 males, with ages ranging from 18 to 65 years old, Spanish as first language, and Argentinian nationality. Participants were recruited in

public places (e.g., a coffee shop, a bus station, a gym) and psychology students were excluded from the sample. Inclusion criteria were similar to the original Spanish sample [16]. Roughly half of the participants ($n = 93$) were young adults (18–30). As to educational level, 101 participants were or had been to college or further. The Spanish data came from a sample that was demographically similar, except for the fact that the educational level was higher (155 subjects were or had been to college or further).

Instruments

Evidence of unidimensionality for the total score was found in the process of constructing and validating the EK scores [16]. This is why the EK test can be described as composed of three subtests (the original tests: Emotion Vocabulary, Close Emotional Situations, Far Emotional Situations).

The test was implemented on a portable computer. Identification, gender, age, consent, response option and right/wrong answers are stored by the application. Each of the three subtests is composed of forty multiple-choice items, eight for each of the five emotion "families". Each item is composed of a stem and five response options: happiness, sadness, anger, fear, and disgust. Fig 1 shows three item examples.

During item construction, two judges, one for each country, evaluated the content seeking to avoid lexical and situational peculiarities (e.g. words having a slang meaning not contained in the dictionary, scenarios reflecting local particularities). Words and scenarios had to represent emotional prototypes equally understood in both countries.

Emotion Vocabulary (EV). The subtest is composed of items 1–40. Each item stem is an emotion word whose frequency per million is similar in Argentina and in Spain according to CORPES XXI [26]. The participant is asked to choose the response option whose meaning is the closest to that of the target word. An EV item example can be seen in Fig 1A.

Close Emotional Situations (CES). The subtest is composed of items 41–80. Item stems are verbal scenarios that show a character and a close/concrete act, object, moment, and place. Scenarios describe concrete variations of the emotion prototypes. The participant is asked to choose the option that best describes the emotion that would be typical to feel in that situation. A CES item example can be seen in Fig 1B.

Far Emotional Situations (FES). The subtest is composed of items 81–90. Item stems are verbal scenarios that show a far/abstract character, time and situation. Scenarios describe abstract variations of the emotion prototypes. The subject is asked to choose the option that best describes the emotion that would usually be felt in that abstract situation. A FES item example can be seen in Fig 1C.

Procedure

A university researcher approached participants individually and asked about age, place of residence and first language (inclusion criteria). Individual privacy and anonymity were protected. Following the usual procedures in psychological research, data was aggregated and participants gave informed consent (the computerized test includes a button "I consent" to start the tasks.) The test was applied on a portable computer; administration took between fifteen and thirty minutes. Subjects were debriefed about the study upon completion of the tasks.

Ethical statement

The participants were treated in accordance with the Helsinki ethical guidelines. The Spanish MINECO responsible committee revised the application (including ethical aspects), and approved the research under Grant PSI2014-52369-P. All participants provided their informed

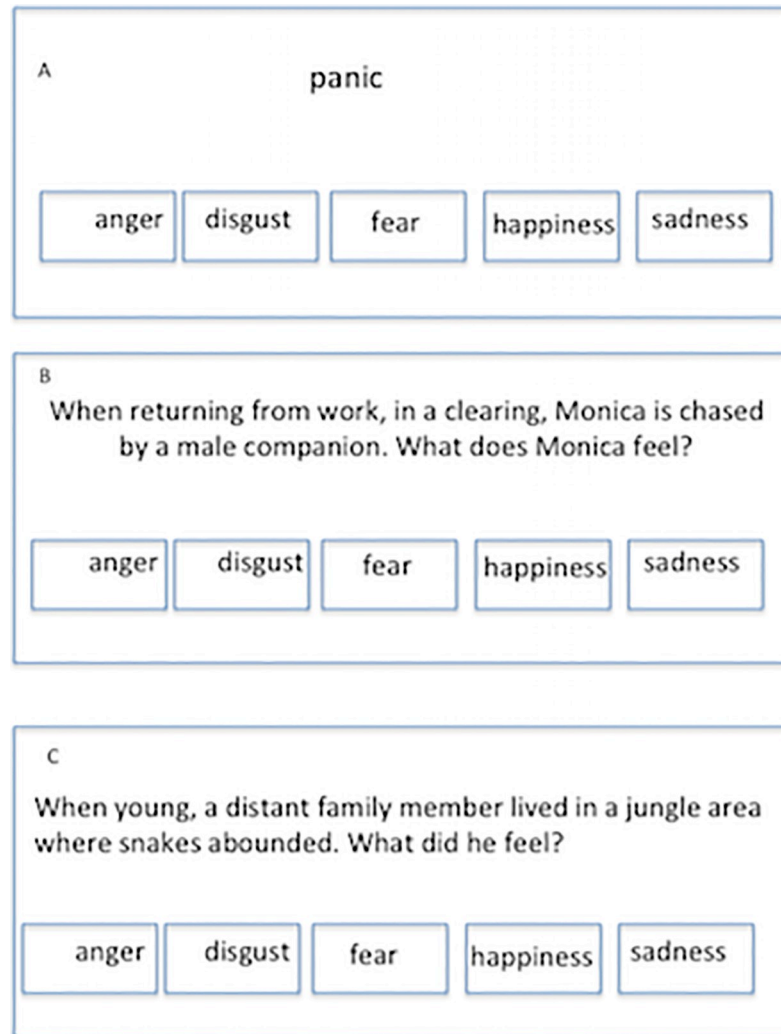


Fig 1. Three FEAR item examples. (A) Emotion Vocabulary item. (B) Close Emotional Situations item. (C) Far Emotional Situations item. Note: Items were written in Spanish, so the translation is an approximation.

<https://doi.org/10.1371/journal.pone.0207335.g001>

consent twice: verbally, while participants were being invited to take part in the study, and via the computer program. Individual privacy and anonymity were protected.

Data analysis

Rasch analyses were performed with Winsteps 3.80.1 [24]. Data-model fit was assessed by means of *infit* (an information-weighted form of *outfit*) and *outfit* (calculated by adding the standardized square of residuals after fitting the model over items or subjects to form chi-square-distributed variables). *Infit/outfit* values over 2 are not adequate for the measurement system [24]. Component analyses of residuals are performed by Winsteps 3.80.1 in order to test the unidimensionality assumption. The recommendations are that Rasch measures should account for at least 20% of the total variance [27] and that the unexplained variance in the first contrast be low [28]. As to the assumption of local independence, it was assessed with Yen's Q3 test [29]. High positive correlation of residuals for two items shows that they may be locally dependent. It is usual to compute the correlation matrix of residuals and select the maximum

Table 1. EK items: Emotion, score, Rasch Di and SE.

001	HAPPINESS	395	-2.26	.45
002	DISGUST	234	1.99	.11
003	ANGER	380	-.79	.23
004	FEAR	393	-1.91	.39
005	DISGUST	392	-1.77	.36
006	DISGUST	319	.87	.13
007	HAPPINESS	398	-3.19	.71
008	SADNESS	365	-.16	.18
009	ANGER	305	1.09	.12
010	DISGUST	260	1.69	.11
011	SADNESS	385	-1.10	.27
012	HAPPINESS	399	-3.89	1.00
013	HAPPINESS	377	-.64	.22
014	SADNESS	364	-.13	.18
015	DISGUST	328	.72	.13
016	FEAR	294	1.25	.12
017	FEAR	325	.77	.13
018	HAPPINESS	400	—	—
019	ANGER	335	.58	.14
020	HAPPINESS	375	-.55	.21
021	FEAR	117	3.33	.11
022	ANGER	290	1.31	.12
023	DISGUST	118	3.32	.11
024	HAPPINESS	277	1.48	.11
025	FEAR	316	.92	.13
026	SADNESS	267	1.60	.11
027	ANGER	358	.05	.17
028	SADNESS	356	.10	.16
029	HAPPINESS	360	-.01	.17
030	DISGUST	245	1.86	.11
031	FEAR	145	2.99	.11
032	FEAR	393	-1.91	.39
033	SADNESS	339	.50	.14
034	DISGUST	92	3.68	.12
035	ANGER	370	-.34	.19
036	SADNESS	326	.75	.13
037	SADNESS	331	.66	.14
038	FEAR	382	-.91	.25
039	ANGER	98	3.60	.12
040	ANGER	292	1.28	.12
041	ANGER	306	1.08	.12
042	FEAR	379	-.74	.23
043	FEAR	388	-1.34	.30
044	FEAR	353	.18	.16
045	HAPPINESS	395	-2.26	.45
046	ANGER	223	2.11	.11
047	SADNESS	296	1.22	.12
048	DISGUST	213	2.22	.11

(Continued)

Table 1. (Continued)

049	HAPPINESS	396	-2.49	.51
050	HAPPINESS	397	-2.78	.58
051	FEAR	378	-.69	.22
052	SADNESS	380	-.79	.23
053	ANGER	333	.62	.14
054	ANGER	325	.77	.13
055	ANGER	354	.16	.16
056	DISGUST	272	1.54	.11
057	HAPPINESS	394	-2.07	.42
058	DISGUST	331	.66	.14
059	SADNESS	313	.97	.13
060	DISGUST	369	-.30	.19
061	ANGER	348	.30	.15
062	HAPPINESS	392	-1.77	.36
063	SADNESS	351	.23	.16
064	FEAR	352	.21	.16
065	FEAR	387	-1.26	.29
066	DISGUST	382	-.91	.25
067	FEAR	393	-1.91	.39
068	ANGER	378	-.69	.22
069	DISGUST	370	-.34	.19
070	SADNESS	341	.46	.15
071	ANGER	381	-.85	.24
072	HAPPINESS	397	-2.78	.58
073	SADNESS	372	-.42	.20
074	FEAR	386	-1.18	.28
075	HAPPINESS	391	-1.65	.34
076	DISGUST	360	-.01	.17
077	SADNESS	245	1.86	.11
078	SADNESS	333	.62	.14
079	HAPPINESS	392	-1.77	.36
080	DISGUST	300	1.17	.12
081	DISGUST	347	.33	.15
082	ANGER	323	.80	.13
083	FEAR	371	-.38	.20
084	SADNESS	279	1.45	.11
085	HAPPINESS	394	-2.07	.42
086	ANGER	327	.73	.13
087	ANGER	331	.66	.14
088	DISGUST	350	.26	.16
089	SADNESS	329	.70	.14
090	HAPPINESS	396	-2.49	.51
091	DISGUST	366	-.20	.18
092	FEAR	341	.46	.15
093	HAPPINESS	393	-1.91	.39
094	HAPPINESS	383	-.97	.25
095	SADNESS	330	.68	.14
096	ANGER	360	-.01	.17

(Continued)

Table 1. (Continued)

097	SADNESS	233	2.00	.11
098	SADNESS	362	-.07	.18
099	DISGUST	286	1.36	.12
100	DISGUST	384	-1.03	.26
101	FEAR	203	2.33	.10
102	FEAR	351	.23	.16
103	SADNESS	373	-.46	.20
104	ANGER	323	.80	.13
105	FEAR	385	-1.10	.27
106	FEAR	351	.23	.16
107	HAPPINESS	393	-1.91	.39
108	DISGUST	317	.90	.13
109	SADNESS	389	-1.44	.31
110	HAPPINESS	390	-1.54	.32
111	ANGER	310	1.02	.12
112	FEAR	386	-1.18	.28
113	HAPPINESS	393	-1.91	.39
114	SADNESS	378	-.69	.22
115	DISGUST	210	2.26	.10
116	HAPPINESS	393	-1.91	.39
117	ANGER	339	.50	.14
118	ANGER	326	.75	.13
119	FEAR	375	-.55	.21
120	DISGUST	382	-.91	.25

<https://doi.org/10.1371/journal.pone.0207335.t001>

value (Q3,max). However, no single stand-alone critical value exists, and the range of residual correlations values is influenced by various factors, including the number of items [30]. In practical terms, correlations over .70 would be clearly indicative of local dependence (Linacre, 2013).

As to DIF, it was analyzed by testing the standardized difference between item calibrations in two groups across three criteria (gender: 0 = female, 1 = male; age: 0 = below college, 1 = college and over; country: 0 = Spain, 1 = Argentina) with Bonferroni-corrected *alpha* levels; the Rasch-modeled scores from the analysis of all the participants were held constant, providing the conjoint measurement scale in *logit* units. *Welch-t* and *Cohen's d* were calculated to test differences between groups on Rasch scores, before and after removing the seven country-related DIF affected items.

Results

One (happiness) item got perfect score and therefore its Rasch measure was not estimated. The Rasch analysis of the remaining data indicates good data-model fit for items, *mean infit* was .99 (*SD* = .05) and *mean outfit* was .90 (*SD* = .21). For persons, *mean infit* was 1.00 (*SD* = .16) and *mean outfit* was .90 (*SD* = .50). No item showed *infit/outfit* over 1.5. Eleven persons (less than 3%) showed *outfit* over 2, but none of them showed *infit* over 2. The percentage of variance explained by EK measures was 24.8% and the component analysis of residuals showed that the unexplained variance in the first contrast was 2.7%. *Item reliability* (.97) and *model person reliability* (.80) were adequate. Residual correlations between items were in the range (-.23,.67), with average 0.00. There were no residual correlations over .70. Less than 3 per 1000

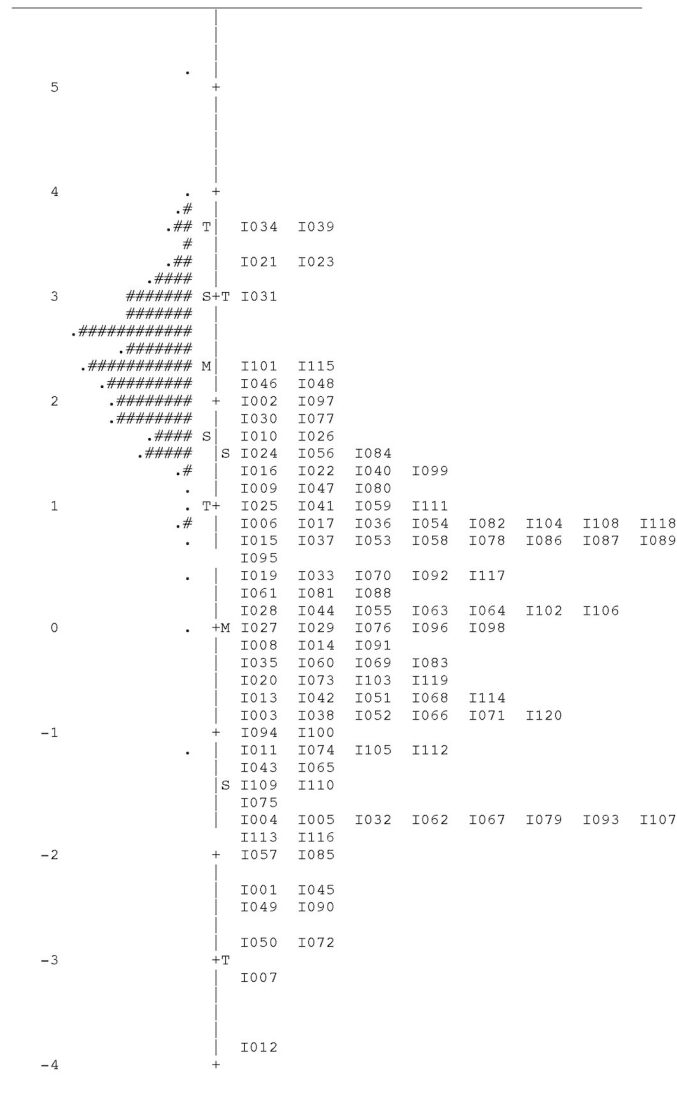


Fig 2. Emotion knowledge: Map of the variable. Note: M = mean; S = 1 SD; T = 2 SD; each "#" is 4; each "." is 1 to 3.

<https://doi.org/10.1371/journal.pone.0207335.g002>

residual correlations were in the range .40-.67. Thus, the assumption of local independence for items can be maintained. Table 1 shows the main results of the item analysis.

The map of the variable (or Wright map) can be seen in Fig 2: person measures are on the left while the right side shows item difficulties.

Average person aptitude in *logit* units was 2.36, *SD* = .68, *range* = -1.12 to 5.09. No item showed sex-related DIF, nor were gender differences (impact) found in Rasch measures, *Welch-t* (385) = 1.96, *p* = .051, *d* = -.19 (conventionally coded as 0 = female, 1 = male).

Five items (I20, I24, I29, I56 and I58) showed age-related DIF; two of them, I56 and I58, favored the young group, and thus DIF can be considered as balanced (i.e., a small number of items favored each of the two groups and so it is considered of no consequence). No age-related differences in Rasch measures were found, *Welch-t* (396) = -1.84, *p* = .067, *d* = .18 (coded as 0 = 18–30, 1 = 31–65). Education level was coded as 0 = below college, 1 = college and over. Two items (I24, which favored the less educated group, and I36) showed education-

related balanced DIF (i.e., one item favored each of the two groups and so it is considered of no consequence); small-sized education-related differences in Rasch measures were found, $Welch-t(355) = -2.69, p = .008, d = .25$.

Seven items (I23, I26, I27, I30, I31, I80 and I101) showed country-related DIF, five of which favored the Spanish participants, and two favored the Argentinian ones (I80 and I101). Small significant differences in Rasch measures were found, $Welch-t(362) = 2.54, p = .011, d = -.25$ (coded as 0 = Spain, 1 = Argentina). Mean scores were 2.45 in Spain and 2.28 in Argentina.

After deleting these seven items, the Rasch analysis of the remaining data showed good fit for items: *mean infit* was .99 ($SD = .06$), and *mean outfit* was .89 ($SD = .22$). For persons, *mean infit* was .99 ($SD = .15$) and *mean outfit* was .89 ($SD = .48$). No item showed *infit/outfit* over 1.5. Twelve persons (3%) showed *outfit* over 2, but none of them showed *infit* over 2.

The percentage of variance explained by EK measures was 22.9% and the component analysis of residuals showed that the unexplained variance in the first contrast was 2.9%. *Item reliability* (.97) and *model person reliability* (.79) were good. Differences in EK scores associated with sex ($Welch-t(387) = 1.86, p = .064, d = -.19$, conventionally coded as 0 = female, 1 = male), age ($Welch-t(397) = -1.50, p = .14, d = .15$), educational level ($Welch-t(353) = -2.19, p = .029, d = .23$) and country ($Welch-t(374) = .89, p = .37, d = -.08$, coded as 0 = Spain, 1 = Argentina) were non-significant (Bonferroni-corrected).

Discussion

This study examined whether the EK test showed DIF in two Spanish speaking countries sharing the same language and showing cultural similarities. Based in the conceptual act theory [4], agreement within broad emotional categories for people belonging to a general culture and language was expected, even though some systematic sub-cultural variation in emotional knowledge could also appear.

The generalized validity of the EK test [16] in Argentina was tested with the RM, an implementation of the invariant measurement approach [20, 21]. Results indicated that both fit to the RM and reliability were adequate. There were no significant sex-related or age-related differences in EK. Small differences were found for educational level and country. However, these differences disappeared when the seven country-related DIF affected items were removed. These results are in agreement with the conceptual act theory predictions of a general absence of DIF between the two countries. Only a few items exhibited DIF, probably reflecting some sub-cultural differences. However, this could also be due to *overfitting*: the tendency for statistical models to mistakenly fit sample-specific noise as if it were signal. Minimizing overfitting is needed when the objective is to generalize to new observations that are similar (but not identical) to the ones that have been sampled [31]. This is why we do not recommend deleting these seven items now. If our results are replicated in future studies, then substitution of the seven items must be considered.

Current evidence is sufficient to allow for the EK test to be employed in both Argentina and Spain, in academic or applied settings where individual differences in emotional competence might be relevant. The map of the variable (or Wright map) makes it easy to communicate test results to both academicians and lay people [32]. However, some limitations of our study must be taken into account: the initial validation of the EK scores was carried out on adult samples without disabilities, and so our conclusion is neither applicable to children nor to populations with special needs as, e.g., deaf people. Increasing the number of difficult items is certainly needed in order to reliably assess EK aptitude in high ability samples. We are currently planning to increase the number of high-difficulty emotional vocabulary items.

Supporting information

S1 File. Data file.
(TXT)

Acknowledgments

We want to thank Alicia Monreal Bartolomé, Monica Luccarelli, Yamila Coccimiglio, and Federico Martín González, who helped with the data collection for this project.

Author Contributions

Conceptualization: Ana R. Delgado.

Data curation: Ana R. Delgado, Gerardo Prieto.

Formal analysis: Gerardo Prieto.

Funding acquisition: Ana R. Delgado.

Investigation: Debora I. Burin.

Methodology: Ana R. Delgado, Debora I. Burin, Gerardo Prieto.

Resources: Ana R. Delgado, Debora I. Burin, Gerardo Prieto.

Supervision: Ana R. Delgado, Debora I. Burin.

Writing – review & editing: Ana R. Delgado, Debora I. Burin, Gerardo Prieto.

References

1. Barrett LF. Solving the emotion paradox: Categorization and the experience of emotion. *Pers Soc Psychol Rev.* 2006; 10: 20–46. https://doi.org/10.1207/s15327957pspr1001_2 PMID: 16430327
2. Barrett LF. Variety is the spice of life: A Psychologist Constructionist approach to understanding variability in emotion. *Cogn Emot.* 2009; 23: 1284–1306. <https://doi.org/10.1080/02699930902985894> PMID: 20221411
3. Barrett LF. Emotions are real. *Emotion* 2012; 12: 413–429. <https://doi.org/10.1037/a0027555> PMID: 22642358
4. Barrett LF. The conceptual act theory: a précis. *Emot Rev.* 2014; 6: 292–297. <https://doi.org/10.1177/1754073914534479>
5. Lindquist KA, Satpute AB, Gendron M. Does language do more than communicate emotion? *Curr Dir Psychol Sci.* 2015; 24: 99–108. <https://doi.org/10.1177/0963721414553440> PMID: 25983400
6. Ekman P, Cordaro D. What Is Meant by Calling Emotions Basic. *Emot Rev* 2011; 3: 2364–2370. <https://doi.org/10.1177/1754073911410740>
7. Fehr B, Russell JA. Concept of emotion viewed from a prototype perspective. *J Exp Psychol Gen.* 1984; 113: 464–486. <https://doi.org/10.1037/0096-3445.113.3.464>
8. Widen SC, Russell JA. Children acquire emotion categories gradually. *Cognitive Dev.* 2008; 23: 291–312. <https://doi.org/10.1016/j.cogdev.2008.01.002>
9. Izard CE, Woodburn EM, Finlon KJ, Krauthamer-Ewing ES, Grossman SR, Seidenfeld A. Emotion knowledge, emotion utilization, and emotion regulation. *Emot Rev* 2011; 3: 44–52. <https://doi.org/10.1177/1754073910380972>
10. Roberts RD, Schulze R, O'Brien K, MacCann C, Reid J, Maul A. Exploring the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT); with established emotions measures. *Emotion* 2006; 6: 663–669. <https://doi.org/10.1037/1528-3542.6.4.663> PMID: 17144757
11. Mayer JD, Salovey P, Caruso DR. The validity of the MSCEIT: Additional analyses and evidence. *Emot Rev.* 2012; 4: 403–408. <https://doi.org/10.1177/1754073912445815>
12. Mayer JD, Caruso DR, Salovey P. The ability model of Emotional Intelligence: principles and updates. *Emot Rev* 2016; 8: 290–300. <https://doi.org/10.1177/1754073916639667>

13. Herpertz S, Nizielski S, Hock, Schüt A. The relevance of Emotional Intelligence in personnel selection for high emotional labor jobs. *PLoS ONE* 11(4): e0154432. <https://doi.org/10.1371/journal.pone.0154432> PMID: 27124201
14. Trentacosta CJ, Fine SE. Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Soc Dev.* 2010; 19: 1–29. <https://doi.org/10.1111/j.1467-9507.2009.00543.x> PMID: 21072259
15. Wojciechowski J, Stolarski M, Matthews G. Emotional intelligence and mismatching expressive and verbal messages: a contribution to detection of deception. *PLoS ONE* 9(3): e92570. <https://doi.org/10.1371/journal.pone.0092570> PMID: 24658500
16. Delgado AR, Prieto G, Burin DI. Constructing three emotion knowledge tests from the invariant measurement approach. *PeerJ* 5:e3755; <https://doi.org/10.7717/peerj.3755> PMID: 28929013
17. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
18. Delgado AR. Using the Rasch Model to quantify the causal effect of instructions. *Behav Res Methods.* 2007; 39: 570–573. PMID: 17958169
19. Delgado AR. Measuring Emotion Understanding with the Rasch Model. *Actual Psicol.* 2016; 30: 47–56. <http://dx.doi.org/10.15517/ap.v29i119.21516>
20. Engelhard G. Invariant measurement: Using Rasch Models in the social, behavioral and health sciences. New York, NY: Routledge; 2013.
21. Engelhard G, Wang J. Alternative measurement paradigms for measuring executive functions: SEM (formative and reflective models) and IRT (Rasch models). *Measurement* 2014; 12: 102–108. <https://doi.org/10.1080/15366367.2014.943592>
22. Prieto G, Delgado AR, Perea MV, Ladera V. Scoring neuropsychological tests using the Rasch Model: An Illustrative Example With the Rey-Osterreith Complex Figure. *Clin Neuropsychol.* 2010; 24: 45–56. <https://doi.org/10.1080/13854040903074645> PMID: 19658034
23. Wu M, Tam HP, Jen TH. Educational measurement for applied researchers: Theory into practice. Singapore: Springer Nature; 2017.
24. Linacre JM. Winsteps Rasch measurement computer program, version 3.80.1. Chicago: [Winsteps.com](http://www.winsteps.com); 2013.
25. Prieto G, Nieto E. Influence of DIF on differences in performance of Italian and Asian individuals on a reading comprehension test of Spanish as a foreign language. *J Appl Meas.* 2014; 15: 176–188. PMID: 24950535
26. Real Academia Española. Corpus del Español del Siglo XXI, CORPES XXI; 2015. [Internet]. Available from: [<http://web.frl.es/CORPES/view/inicioExterno.view>]
27. Reckase MD. Unifactor latent trait models applied to multifactor tests: results and implications. *J Educ Stat.* 1979; 4: 207–230.
28. Miguel JP, Silva JT, Prieto G. Career Decision Self-Efficacy Scale-Short Form: A Rasch analysis of the Portuguese version. *J Vocat Behav* 2013; 82: 116–123. <https://doi.org/10.1016/j.jvb.2012.12.001>
29. Yen WM. Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *App Psychol Meas* 1984; 8: 125–145. <https://doi.org/10.1177/014662168400800201>
30. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of local dependence in the Rasch Model using residual correlations. *App Psychol Meas* 2017; 41: 178–194. <https://doi.org/10.1177/0146621616677520> PMID: 29881087
31. Tarkoni T, Westfall J. Choosing prediction over explanation in psychology: Lessons From Machine Learning. *Perspects Psychol Sci* 2017; 12: 1100–1122. <https://doi.org/10.1177/1745691617693393> PMID: 28841086
32. Wilson M, Draney K. A technique for setting standards and maintaining them over time. In Nishisato S, Baba Y, Bozdogan H, and K. Kanefuji K, editors. *Measurement and Multivariate Analysis*. Tokyo: Springer Japan. 2002; pp. 325–332.