Pontifícia Universidade Católica
do Rio de Janeiro

Luis Flavio Chaves Anunciação

Use of the Ages and Stages Questionnaires to establish indicators of child development in public daycare centers in Rio de Janeiro, Brazil.

Tese de doutorado

Thesis presented to the Programa de Pós-graduação em Psicologia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Psicologia.

Advisor: Prof. Jesus Landeira-Fernandez

Rio de Janeiro
March 2018

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Luis Flavio Chaves Anunciação

Use of the Ages and Stages Questionnaires to establish indicators of child development in public daycare centers in Rio de Janeiro, Brazil

Thesis presented to the Programa de Pós-graduação em Psicologia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Psicologia. Approved by the undersigned Examination Committee.

Prof. Jesus Landeira-Fernandez
Advisor
Departamento de Psicologia - PUC-Rio

Prof. Cristiano Augusto Coelho Fernandes
Departamento de Engenharia Elétrica - PUC-Rio

Prof. Marcelo Cunha Medeiros
Departamento de Economia - PUC-Rio

Prof. Daniel Domingues dos Santos
Faculdade de Economia - USP

Prof. Pedro Paulo Pires dos Santos
Instituto de Psicologia - UFF

Prof. Bruno Figueiredo Damásio
Instituto de Psicologia - UFRJ

Profa. Monah Winograd
Head of Graduate Studies
Centro de Teologia e Ciências Humanas – PUC-Rio

Rio de Janeiro, March 2nd, 2018

Luis Flávio Chaves Anunciação

Graduated in Psychology at Instituto Brasileiro de Medicina de Reabilitação (Brazil) and obtained his M.Sc. Degree in Public Health from University of State of Rio de Janeiro (UERJ) in 2012. Currently, I work with psychometrics and statistics.

Bibliographic data

Anunciação, Luis Flávio Chaves

Use of the ages and stages questionnaires to establish indicators of child development in public daycare centers in Rio de Janeiro, Brazil / Luis Flávio Chaves Anunciação ; advisor: Jesus Landeira-Fernandez. – 2018.

126 f. : il. color. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Psicologia, 2018.

Inclui bibliografia

1. Psicologia – Teses. 2. Psicometria. 3. Avaliação psicológica. 4. ASQ. 5. Desenvolvimento infantil. 6. Políticas públicas. I. Landeira-Fernandez, J. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Psicologia. III. Título.

CDD: 150

Acknowledgments

I am grateful to Professor Landeira-Fernandez and Cristiano Fernandes for all support they provided during my doctoral study. I am also grateful to Professor Jane Squires for having provided support while I was in the United States.

My title is a somewhat free paraphrase of the expressed aim of the Psychometric Society, namely, to further "*the development of psychology as a quantitative rational science*" (Thurstone, 1937). I think we are all agreed that psychology has to do with behavior, that a quantitative discipline involves numbers, and that a rational science requires broad generalizations.
The premises underlying psychometrics seem to be the following:

1. That behavior can be represented numerically;
2. That the resultant numbers can be manipulated according to mathematical principles; and
3. That the results of these manipulations can yield new and valid representations of behavior.

These assumptions are made implicitly every time we quantify a set of observations, compute a descriptive statistic, and generalize to a population of cases not yet encountered. They are so much part and parcel of the psychometrician's kit of tools that only on rare occasions do we bother to scrutinize them.

Philip H. DuBois. Presidential address delivered at the meeting of the Psychometric Society (1962)

Abstract

Anunciação, Luis; Landeira-Fernandez, Jesus (Advisor). **Use of the Ages and Stages Questionnaires to establish indicators of child development in public daycare centers in Rio de Janeiro, Brazil**. Rio de Janeiro, 2018. 126p. Tese de Doutorado – Departamento de Psicologia. Pontifícia Universidade Católica do Rio de Janeiro.

From 2010 to 2012, the city of Rio de Janeiro initiated a large-scale project to evaluate the development of children who were enrolled in daycare centers and public preschools. The evaluation of child development was performed annually, with an average of 60,000 children in ~80% of the daycare centers and public preschools in the city. After receiving specific training, professionals at each institution (e.g., teachers and caregivers) applied the Ages and Stages Questionnaires (ASQ-3 and ASQ:SE) to screen for risk factors (e.g., developmental delays) with regard to the major markers and milestones of child development. However, in the Brazilian context, there is a lack of studies that have investigated the psychometric properties of the Brazilian version of the ASQ:SE. Moreover, a dearth of studies has discussed the results of both the ASQ:BR and ASQ:SE with regard to longitudinal assessments, gender differences, and incorporation of the statistical results into public policies. That said, the present thesis had the following objectives: (1) to investigate the main psychometric properties of the ASQ:SE using both Classical Test Theory (CTT) methodology and Item Response Theory (IRT) and (2) to present an overview of child development in daycare services using the data that were gathered from this longitudinal project. Thus, this thesis comprises five articles in which the following studies were conducted. Initially, based on CTT, the factor structure of the ASQ:SE was estimated using Cronbach's α, and preliminary cutoff points were computed, based on means and standard deviations. The second article extended investigation of the factor structure of the ASQ:SE and in the third article, a multidimensional IRT model was used for two purposes: (*i*) to confirm the dimensionality of all scales of the ASQ:SE and (*ii*) to verify its item characteristics. In the fourth and fifth article, the focus partially shifted from psychometric properties to the preliminary results of a longitudinal evaluation of children who were assessed over two or three consecutive years. The overall results provide evidence of the validity, reliability, and utility of the ASQ:SE and suggest

that children who are enrolled in Brazilian/Rio de Janeiro's public daycare centers, despite environmental adversity, are not at risk of serious developmental delays.

Keywords

Psychometrics; Psychological Evaluation; ASQ; Child development: Public Policy

Resumo

Anunciação, Luis; Landeira-Fernandez, Jesus (Orientador). **Uso do Ages and Stages Questionnaires para o desenvolvimento de indicadores para monitoramento de creches na cidade do Rio de Janeiro, Brasil**. Rio de Janeiro, 2018. 126p. Tese de Doutorado – Departamento de Psicologia. Pontifícia Universidade Católica do Rio de Janeiro.

De 2010 a 2012, a cidade do Rio de Janeiro iniciou um projeto de grande escala para avaliar o desenvolvimento de crianças matriculadas em creches e pré-escolas públicas. A avaliação do desenvolvimento infantil foi realizada anualmente, com uma média de 60 mil crianças em cerca de 80% das creches e pré-escolas públicas da cidade. Depois de receber treinamento específico, os profissionais de cada instituição (por exemplo, professoras e crecheiras) aplicaram os questionários que compõe o sistema *Ages and Stages Questionnaire* (ASQ-3 e ASQ:SE) para detectar possíveis condições de risco em relação aos principais marcadores do desenvolvimento infantil. Entretanto, no contexto brasileiro, há uma ausência de estudos que tenham investigado as propriedades psicométricas da versão brasileira do ASQ:SE. Além disso, poucos estudos discutiram os resultados do ASQ:BR e ASQ:SE com respeito a avaliações longitudinais, diferenças de performance em função do gênero e visando incorporação dos resultados estatísticos em políticas públicas. Dito isto, a presente tese teve os seguintes objetivos: (1) investigar as principais propriedades psicométricas do ASQ:SE usando metodologia de Teoria Clássica dos Testes (TCT) e da Teoria da Resposta do Item (TRI) e (2) apresentar uma visão geral do desenvolvimento infantil em serviços de creche usando, para isso, os dados que foram coletados longitudinalmente por este projeto. Assim, esta tese compreende cinco artigos nos quais foram realizados os seguintes estudos. Inicialmente, com base na TCT, a estrutura fatorial do ASQ:SE foi investigada, a fidedignidade foi calculada usando o alfa de Cronbach, e pontos de corte preliminares foram calculados com base em médias e desvios-padrão. O segundo artigo ampliou a investigação da estrutura de fatorial do ASQ:SE e no terceiro artigo, um modelo de multidimensional da TRI foi usado para duas finalidades: (i) confirmar a dimensionalidade de todas as escalas do ASQ:SE e (ii) verificar as características de seus itens. No quarto e quinto artigo, o foco foi parcialmente

alterado das propriedades psicométricas para os resultados preliminares de uma avaliação longitudinal de crianças que foram avaliadas ao longo de dois ou três anos consecutivos. Os resultados gerais evidenciam a validade, confiabilidade e utilidade do ASQ:SE e sugerem que crianças que estão matriculadas em creches públicas do Brasil/Rio de Janeiro, apesar da adversidade ambiental, não estão em risco de atrasos sérios no desenvolvimento.

Palavras chave

Psicometria; Avaliação Psicológica; ASQ; Desenvolvimento infantil, Políticas Públicas

Table of content

List of Figures

List of Tables

I. Introduction

From 2010 to 2012, the city of Rio de Janeiro initiated a large-scale project to evaluate the development of children who were enrolled in daycare centers and public preschools. The goal was to develop an evidence-based public policy monitoring system that can map the needs of different communities, social groups, and children (Barros, Coutinho, & Mendonça, 2016).

The project was included as one part of existing strategies that had extensive federal support to evaluate the effectiveness of the Brazilian public education system. In conjunction with adolescents' scores on the National High School Exam (*ENEM*), a cognitive measure would allow evaluation of the quality of high schools to further improve the educational system (BRASIL, 2011). This initiative was influenced by the notion of accountability because all data would enhance transparency in public management (Fernald, Prado, Kariger, & Raikes, 2017).

The evaluation of child development was performed annually, with an average of 60,000 children in ~80% of the daycare centers and public preschools in the city (Costin, n.d.). After receiving specific training, professionals at each institution (e.g., teachers and caregivers) applied the Ages and Stages Questionnaires (ASQ-3 and ASQ:SE) to screen for risk factors (e.g., developmental delays) with regard to the major markers and milestones of child development.

The ASQ and ASQ:SE are composed of Likert scales. The respondent reads a question about a specific/target behavior (e.g., "Does the child look at you when you talk to him?") and marks the frequency of this behavior: "always," "sometimes," or "rarely." Based on statistical analysis, the results represent the following developmental domains: communication, gross motor, fine motor, problem-solving, and personal and social skills. The study that validated the ASQ obtained data from 2,008 children to determine its psychometric qualities (Squires, Bricker, & Potter, 1997). The study that validated the ASQ:SE was conducted with 3,014 children (Squires, Bricker, Heo, & Twombly, 2001).

The ASQ is based on three important premises. First, at-risk infants and children do not necessarily present delays or disabilities at the same point in time (e.g., at 6 months). Thus, screening should be dynamic and designed to follow a child's development over time rather than rely on testing at one or two age intervals.

Second, individuals who have the greatest familiarity with the child should be involved in the screening process. Parents and other caregivers often know more about their children than even the most sophisticated assessment specialist. Third, for broad implementation, the screening of individual children needs to be low-cost. The inclusion of parents in the screening process affords enormous savings if they can accurately assess their child (Bricker et al., 2008).

Thus, by enabling a simple, comprehensive screening assessment and because of its psychometric properties, the ASQ and ASQ:SE have been adapted to different contexts. Validation studies have been conducted for use of the ASQ in Canada, Mexico, Portugal, Spain, France, Netherlands, Norway, Turkey, Chile, Portugal, and Brazil (A. Singh, Squires, Yeh, Heo, & Bian, 2016).

However, in the Brazilian context, there is a lack of studies that have investigated the psychometric properties of the Brazilian version of the ASQ:SE. Moreover, a dearth of studies has discussed the results of both the ASQ:BR and ASQ:SE with regard to longitudinal assessments, gender differences, and incorporation of the statistical results into public policies.

2. Theoretical Background

Psychometric background

When one decides to use a psychological instrument such as a questionnaire or a test, the decision comes with an inherent understanding and agreement that psychological characteristics, traits or abilities can be investigated in a systematic manner. Another agreement is made when one decides to analyze the data obtained by some tool by summing up the scores or by using other mathematical methods. This latter attitude comes with a deep epistemological acceptance that psychological traits can be casted in numerical form for the underlying structure. Although these premises were already well known and documented in publications by the first psychologists, this paradigm was not entirely accepted by the scientific community until recently.

The close relation between statistics and psychology is well documented, and with the formation of the Psychometric Society in 1935 by L.L. Thurstone, psychometrics is seen as a separate science that interfaces with mathematics and psychology. In a broad sense, the area is concerned with quantifying and analyzing human differences, and in a narrower sense it is concerned with evaluating the attributes of psychological tests and other measures used to assess variability in behavior and then to link such variability to psychological phenomena and theoretical frameworks (Browne, 2000; Furr & Bacharach, 2008).

The two disciplines of psychometrics and statistics have at least three points in common. First, they use models to simplify and study the reality; second, they are highly dependent on mathematics; and third, both can be observed by its tools (e.g. statistical inference tests are provided by statistics and/or psychological instruments are provided by psychometrics) or by their theoretical framework,

where researchers seek to build new models and paradigms through guidelines, empirical data and simulations.

Strictly speaking, psychological phenomena such as attention and extraversion are not directly observable, nor can they be measured directly. Because of that, they must be inferred from observations made on some behavior that may be observed and is assumed to operationally represent the unobservable characteristic (or "variable") that is of interest. There are numerous synonyms in the literature when referring to non-directly observable psychological phenomena such as abilities, constructs, attributes, latent variables, factors or dimensions (Furr & Bacharach, 2008).

There are several avenues available when trying to assess psychological phenomena. Multimethod assessments such as interviews, direct observation, and self-reporting, as well as quantitative tools such as tests and scales are accessible to psychologists (Hilsenroth, Segal, & Hersen, 2003). However, from this group of methods the use of tests, inventories, scales, and other quantitative tools are seen as the best choices when one needs to accurately measure psychological traits (Borsboom, Mellenbergh, & van Heerden, 2003; Craig, 2017; Marsman et al., 2018; Novick, 1980), as long as they are psychometrically adequate.

In line with this, the use of quantitative methods in psychology (and social sciences in general) has been increasing dramatically in the last decades – their use began in the 1960s; but exploded in the 80s), despite strong criticism and concern from different groups that disagree with this quantitative view (Cousineau, 2007). Paradoxically, this quantitative trend was only partially followed by academics and other students of psychology, which has led to the American Psychological

Association creating a task force aiming to increase the number of quantitative psychologists and to improve the quantitative training among students.

With that being said, the aim of this article is to provide a succinct review of the history of psychometrics and its methods through important points of psychometrics. It is important to clarify that this review is not about examining all trends in psychometrics so that it is not exhaustive and has concentrated on describing and summarizing the topics related to this thesis. Several other resources are relevant to the topic and some are listed in the references.

History of Psychometrics

The precise historical origins of psychometrics and the field of quantitative psychology are difficult to define. The same condition is found in statistics when trying to detail when statistics was incorporated into social sciences/humanities. However, it is possible to argue that the investigation into psychometrics has two starting points. The first one was concerned with discovering general laws relating the physical world to observable behavior, and the second one had the aim to explore and to test some hypotheses about the nature of individual differences by using psychological testing (Craig, 2017; Furr & Bacharach, 2008). When arranging events in their order of occurrence in time, James Cattell was the first psychologist to write about psychometrics in 1886 with a thesis entitled "Psychometric Investigation". At this time, Cattel was Wundt's student, but he was highly influenced by Francis Galton and his "Anthropometric Laboratory" which opened in London in 1884. As consequence of the interface between the two researchers, Cattell is also credited as the founder of the first laboratory developed to study psychometrics, which was established within the Cavendish Physics

Laboratory at the University of Cambridge in 1887 (J. M. Cattell, 1928; Ferguson, 1990).

With this first laboratory, the field of psychometrics could differentiate from psychophysics and the major differences can be grouped as the following: 1) while psychophysics aimed to discover general sensory-perception laws (i.e. psychophysical functions), psychometrics was (is) concerned with studying differences between individuals; 2) the goal of psychophysics is to explore the fundamental relations of dependency between a physical stimulus and its psychological response, but the goal of psychometrics is to measure what we call latent variables, such as intelligence, attitudes, beliefs and personality; 3) the methods in psychophysics are based on experimental design where the same subject is observed over repeated conditions in a controlled experiment, but the majority of studies in psychometrics is observational when the measurement occurs without trying to affect the participants (L. V. Jones & Thissen, 2007).

Nowadays, graduate programs in Psychometrics are found in countries such as the United States and division 5 (Quantitative and Qualitative Methods) from the American Psychological Association (APA) helps in studying measurement, statistics, and psychometrics. The Classical Test Theory (CTT) and the Item Response Theory (IRT) are the primary measurement theories employed by researchers in order to construct measures of latent traits, and will be described in the following section.

Classical Test Theory (CTT) and Item Response Theory (IRT)

As previously mentioned, evaluating the attributes of psychological testing is one of the greatest concerns of psychometrics. Different from other sciences such as physics and biology, there are several instruments to measure psychological phenomena. The indirect nature of the instruments leaves much room for unknown sources of variance to contribute to participant's results, which translates into a large measurement error and the conclusion that assessing the validity and the reliability of the psychometric instruments is vital (Peters, 2014). The data yielded by those tests are often used to inform important decisions, including awarding credentials, judging the effectiveness of interventions and making personnel or business decisions. Because of this, it is necessary to ensure the validity and the reliability of an instrument, and various analytic strategies are available to check the properties of an instrument.

Validity is a broad concept and has been widely debated since it was originated in the 1920s. Some authors understand validity as a test property and others consider validity in relation to the test score. The first group state that a test is valid if it measures what it purports to measure (i.e. accuracy), and the second group defines validity as the degree to which evidence and theory support the interpretations of test scores for the proposed uses of the test. The reliability of an instrument represents the consistency, and the reproducibility of its results across different test situations. Reliability quantifies the measurement error and it is expressed in terms of stability, equivalence, and consistency (AERA, APA, & NCME, 2014; Borsboom, Mellenbergh, & van Heerden, 2004; Sijtsma, 2013).

It is possible to recognize two distinct approaches in psychometrics to study these issues and some taxonomies are reported in the literature, which suggest an

absence of consensus among psychometricians. From a historical/traditional standpoint, the "Classical Test Theory" (CTT) and "Item Response Theory" (IRT) are widely used. From a statistical/methodological standpoint, it is not always possible to assign some technique to a particular traditional approach, as it is possible to combine approaches or understand some methods as particular cases of a general approach, such as with Confirmatory Factor Analysis (Edwards & Bagozzi, 2000; Mellenbergh, 1994).

CTT considers that the observed test score is composed of a True score (T) plus an Error (E) considered normally distributed with its mean taken to be 0, $Y = T + E$. It approaches to assume many types of validity (e.g. content, criterion and construct), in which the differentiations may have to do with the validation purpose or with the validation process by specific data analysis techniques. In the last decades, however, there was a growing consensus to take into account validity as a unitary concept, and that differentiations in types of validity should only be considered as different types of gathering evidence (AERA et al., 2014; Sijtsma, 2013).

In regards to the statistical process to explore the constructs covered in psychometric work, there are two main ways in which this connection between constructs and observations has been construed. The first approach understands constructs as inductive summaries of attributes or behaviors as a function of the observed variables (i.e. formative model, where latent variables are formed by their indicators). The second approach understands constructs as reflexive and the presence of the construct is assumed to be the common cause of the observed variables. The statistical model most often associated with this idea is the Principal Component Analysis (PCA) model, and the two major statistical models associated

with the reflective model are the common factor model and the Item Response Theory (IRT) (Edwards & Bagozzi, 2000; Marsman et al., 2018). Image 1 below displays these conceptualizations.

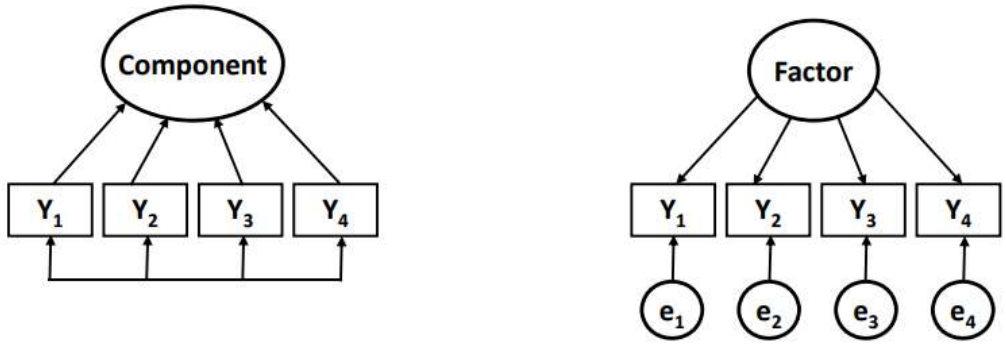

Figure 1 On the left, the PCA model; on the right, the Factor model.

As the goal of PCA is data reduction, but psychometric theory wants to investigate how observable variables are related to theoretical/latent constructs, the reflexive model is mostly used. The Factor Analysis (FA) is part of its models, its concept is analogous to CTT, and was developed with the work of Charles Spearman (1904) in the context of intelligence testing. The FA operates on the notion that measurable and observable variables can be reduced to fewer latent variables that share a common variance and are unobservable (Borsboom et al., 2003). In other words, the statistical purpose of factor analysis is to explain relations among a large set of observed variables using a small number of latent/unobserved variables called factors. FA can be divided into exploratory and confirmatory, and in a broad sense is viewed as a special case of Structural Equation Modeling (SEM) (Gunzler & Morris, 2015).

Exploratory factor analysis (EFA) explores data to determine the number or nature of factors that account for the covariation between variables if the researcher does not own sufficient a priori evidence to establish a hypothesis regarding the number of factors underlying the data. In detail, since there is not an a priori hypothesis about how indicators are related to the underlying factors, EFA is not generally considered a member of the SEM family. In contrast, confirmatory factor analysis (CFA) is a theory-driven model and aims to see whether a particular set of factors can account for the correlations by imposing lower triangular constraints on the factor loading matrix, thus rendering identifiability to the established parameters of the model. In other words, CFA is designed to evaluate the a priori factor structure specified by researchers (T. Brown, 2015; Finch, 2011).

In another direction, some authors argue that there is no clear EFA-CFA distinction in most factor analysis applications, and they fall on a continuum running from exploration to confirmation. Because of this, they choose to call both techniques at a statistics standpoint; an unrestricted model for EFA and a restricted model for CFA. An unrestricted solution does not restrict the factor space, so unrestricted solutions can be obtained by a rotation of an arbitrary orthogonal solution, and all the unrestricted solutions will yield the same fit for the same data. On the other hand, a restricted solution imposes restrictions on the whole factor space and cannot be obtained by a rotation of an unrestricted solution (Ferrando & Lorenzo-Seva, 2000).

Leaving aside these particular questions, several high-quality resources on best practices in EFA and CFA are available, and despite some changes in the mathematical notation or formula, the common factor model is a linear regression

model with observed variables as outcomes (dependent variables) and factors as predictors (independent variable) (See equation 1):

$$Y_i = \left(\sum_{m=1}^{M} \lambda_{im} \eta_m \right) + \epsilon_i$$

Equation 1. Common factor model

Where Y_i is the i*th* observed variable (item score) from a set of I observed variables, η_m is the m*th* of M common factors, $\lambda_{im}$is the regression coefficient (slope, also known as factor loading) relating factor m to Y_i, and ϵ_i is the error term unique for each Y_i. The variance of ε for variable i is known as the variable's uniqueness, whereas 1 – VAR(ϵ) is that variable's communality. This latter concept is equivalent to the regression R^2 and describes the proportion of variability in the observed variable explained by the common factors. In some guidelines, the inclusion of the item intercept μ_i is made, but this parameter usually does not contribute to the covariance matrix (Furr & Bacharach, 2008).

Operationally, some assumptions must be fulfilled before an EFA, such as the proportion of variance among variables that might be common variance, and that the dependent variable covariance matrices are not equal across the levels of the independent variables. The first assumption is tested by the Kaiser-Meyer-Olkin (KMO) test, and the second with the Bartlett test. KMO values between 0.8 and 1 indicate the data is adequate for FA, and a significant Bartlett's test ($p < .05$) means that data matrix is not an identity matrix, which prevents factor analysis from working.

Next, three main questions arise when conducting an EFA: 1. The method of factor extraction; 2. How many factors to settle on for a confirmatory step; and 3. Which factor rotation should be employed. All questions need to be answered by the researcher. The extraction methods reflect the analyst's assumptions about the obtained factors. Their mathematical conceptualization is also based on manipulations of the correlation matrix to be analyzed. There are a number of factors to retain changes throughout the literature and there are many rules of thumb to guide the decision. Finally, all results are often adjusted to become more interpretable.

In summary, the factor extraction methods are statistical algorithms used to estimate loadings, and are composed of techniques such as the minimum residual method, principal axis factoring, weighted least squares, generalized least squares and maximum likelihood factor analysis. The decision of how many factors will be retained relies on many recommendations such as: 1. The rule of an eigenvalue of ≥ 1; 2. The point in a scree plot where the slope of the curve is clearly leveling off; or 3. The interpretability of the factors. It is easy to recognize that these guides can provide contradictory answers and illustrate some degree of arbitrary decisions during this process (Nowakowska, 1983). The factor rotations are classified as either orthogonal, in which the factors are constrained to be uncorrelated (e.g. Varimax, Quartimax, Equamax), or oblique (e.g. Oblimin, Promax, Quartimin) in which this constraint is not present (Finch, 2011).

Another approach in psychometrics independent of the factor analysis developments and apart from CTT is the IRT. The focus of IRT modeling is on the relation between each item response and individual latent trait values, represented by an item response function (IRF, also known as Item Characteristic Curve). In

other words, IRT is focused on the relationship between observed responses to items and the underlying dimension or construct, and assumes a relationship between responses to items and the underlying or latent dimension being assessed by the scale.

Considering a sample of *n* individuals that answered *I* items. *s* = 1, …, *n* and *i* = 1, ..., *I*. Let Y_{ij} be random variables associated with the response of individual *s* to an item *i*. These responses can be dichotomous (e.g. fail or pass) or polytomous (e.g. agree, partially agree, neutral). Let Ω_Y denote the set of possible values of the Y_{ij}, assumed to be identical for each item in the test, and θ_s denotes the latent trait for an individual *s*, and η_i a set of parameters that will be used to model item features. The IRT models arise from different sets of possible responses Ω_Y and different functional forms assumed to describe the probabilities with which the Y_{ij} assume those values, as expressed below (Le, 2014; Sijtsma & Junker, 2006; Zumbo & Hubley, 2017):

$$P\left(Y_{ij} = y \middle| \theta_s, \eta_i\right) = f(y|\theta_s, \eta_i); y \in \Omega_Y$$

Equation 2. General formula of IRT models

The η_i represents the item parameters and may include four distinct types of parameters: parameter "a_i" denotes the discrimination, "b_i" the difficulty, "c_i" the guessing, and "d_i" expresses the probability of a high-ability participant failing to answer an item correctly. The common 4PL model is:

$$P(Y_{is} = 1 \mid \theta_s, a, b, c, d) = c_i + (d_i - c_i)\frac{e^{[a_i(\theta_s - b_i)]}}{1 + e^{[a_i(\theta_s - b_i)]}}, a_i > 0\ ; 0 \leq c \leq 1$$

Equation 3. 4PL IRT model

Which leads to:

$$P(Y_{is} = 1 \mid \theta_s, a, b, c, d) = c_i + (d_i - c_i)\frac{1}{1 + e^{[-a_i(\theta_s - b_i)]}}$$

Equation 4. 4PL IRT model

As can be seen from the equations, there is a conceptual bridge between IRT and logistic regression, but the explanatory (independent) variable in IRT is a latent variable as opposed to an observed variable in logistic regression. In the IRT case, the model will recognize the person's variability on the dimension measured in common by the items and individual differences θ may be estimated (Wu & Zumbo, 2007).

In the origins of IRT, some assumptions (such as unidimensionality and local independence) were held, but IRT models can currently deal with multidimensional latent structure (MIRT) and local dependence. In MIRT, an Item Characteristic Surface (ICS) represents the probability that an examinee with a given ability (θ_s) composite will correctly answer an item. To deal with local independence, Item Splitting is a way for the estimation of item and person parameters (Olsbjerg & Christensen, 2015). In the same direction, the comparison between unidimensional and multidimensional models have shown that as the number of latent traits underlying item performance increase, item and ability parameters estimated under MIRT have less error scores and reach more precise measurement (Kose & Demirtasli, 2012).

As previously stated, the reliability of an instrument is investigated along with the validity during a psychometric examination of an instrument, and it can be

performed via methods within the CTT and IRT framework. As in CTT, the observed test score (Y) is composed of a True score (T) plus an Error (E), the variance is $\sigma_Y^2 = \sigma_T^2 + \sigma_E^2$ and the reliability is $\rho_{(YT)}^2 = \frac{\sigma_T^2}{\sigma_Y^2}$, which is equal to $\rho_{(YT)}^2 = \frac{\sigma_T^2}{\sigma_T^2+\sigma_E^2}$. In this case, reliability is theoretically the correlation between a test-score and the true score squared. The main difference in the concept of reliability with regards to CTT and IRT is the Standard Error (SE) used in the calculation of the reliability coefficient, and some approaches can be found in IRT frameworks such as Fisher Information (Lee, Wallace, Raznahan, Clasen, & Giedd, 2014).

In order to investigate the Goodness-of-fit (GoF) between the measurement models and the obtained data, several model fit indices with their criteria are suggested, including the Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Normed Fit Index (NFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). Each index has its own criteria, however, values equal to or greater than 0.9 are frequently reported for GFI, AGFI, NFI, TLI, and TLI as the baseline to conclude in favor of a good fit. For RMSEA, values should be equal to or lower than 0.08 in order to accept a model (Maydeu-Olivares, 2013).

Both CTT and IRT methods are currently seen as complementary and are frequently used to assess the test validity and respond to other research questions.

Conclusions

The investigator often needs to simplify some representation of reality in order to achieve an understanding of the dominant aspects of the system under study. This is no different in Psychology; models are built and their study allows

researchers to answer well-posed and focused questions. When models are useful, their predictions are analogous to the real world.

This section aimed to explore some aspects of the history of psychometrics and to describe its main models. Because the use of psychometrics tools is becoming an important part of several sciences, understanding the concepts presented in this paper will be of importance to mainly enhance the abilities of social and educational researchers.

Child development and the ASQ system

Child development is part of human development which covers the entire lifespan from conception to death, and refers to changes in physical, cognitive, social, emotional and functioning dimensions over time, from conception through adolescence. The changes include physical alterations in size, shape and function, and these alterations can be either progressive or regressive, occur in orderly ways and last for a reasonable period of time (GPH Panel of Experts, 2018).

Child development is seen through different viewpoints depending on theoretical framework and researchers often disagree on critical issues. On the other hand, consensus has emerged on several important points, including: 1. Child development is influenced by the interaction between genetic and environmental factors; 2. It is age-related; 3. It has universal changes; 4. It occurs across a number of interconnected domains; and 5. Infancy and early experience are fundamental to later development/outcomes.

The first assumption means that the debate between the nativist or nurturist views are better understood in terms of the interaction. Second, when one considers

that development is age related, that is not the same as to say that development is age determined; the relationship with age is not seen as deterministic and children learn and mature at different rates. Third, there is a universal pattern in development, but individual differences can account for/explain the variability of the progress rate, as previously stated. Fourth, the development in each domain is closely interwoven with development in the others, although it may not proceed evenly across domains in a parallel fashion (e.g. language development may at times surpass emotional development or vice versa); and finally, as the early experiences are fundamental in building a foundation for cognitive development, deprived childhood may well have grave consequences (Papalia, Olds, & Feldman, 2008).

In summary, every child first communicates with sounds and gestures, and then he/she learns words and how to use them to express her/himself. However, these series of behaviors are subject to a wide range of genetic, environmental and individual influences, which leads to the conclusion that child development (and developmental change) is a dynamic process. In other words, even though there is a pattern of development which is common to all individuals, no two children are alike. They can vary in aspects like preferences, skills, and abilities, etc.

Emotional and social development are seen as competencies formed by non-cognitive skills, knowledge, attitudes, and dispositions necessary to set goals, manage behaviors, build relationships, and process and remember information within settings that can intentionally nurture these competencies. Different terms are used throughout literature, but they have similar definitions or similar cores (S. M. Jones & Kahn, 2017).

Emotional development refers to the emergence of emotions like joy, happiness and fear. Social development refers to the development of abilities that enables the individual to behave in accordance with the expectation of the society. The two processes can be seen as one connected/integrated process during early childhood (or at least the effect of each is on another, which cannot be separated out easily), and at some point in development both processes begin to differentiate from each other (GPH Panel of Experts, 2018).

With that being said, it is well established in the literature nowadays that non-cognitive skills are decisive for children's success. Several research studies suggest that early emotional and social competencies facilitate children's successful interactions with others and are linked to later academic achievement, better overall and mental health, less criminal activity, and less substance use (Jee et al., 2010).

Indeed, studies on brain structure indicate that emotional and social processes are inseparable from cognitive processes. More importantly, from the neuroscience perspective emotional processes help us prioritize, decide, anticipate and plan, while non-cognitive skills are decisive for children's success and the ability to develop their emotional and social competences, which are apparently more influential than cognitive abilities for personal, career and scholastic success (Durlak & Wells, 1997).

Consequently, the measurement of emotional and social development can be a powerful way to achieve large-scale and lasting results for individuals and communities, and therefore is strategic for public policies (Rio de Janeiro City, 2017).

The Ages and Stages Questionnaires Social-Emotional (ASQ:SE) is part of a wider screening instrument for developmental delays called the Ages and Stages

Questionnaires (ASQ). The ASQ:SE began its development in 1995 and was created to capture a valid picture of children's emotional and social developmental status. These tools were designed to assess children at multiple points (6, 12, 18, 24, 30, 36, 48, and 60 months old) in a concise and a simple manner.

The theoretical orientation of ASQ:SE is developmental and its items were designed to target only concrete and objectively definable responses that could be readily observed (e.g. walking, saying words). Validity and reliability studies were conducted on the ASQ:SE between 1996 and 2001 with a normative sample of 3,014 children. Internal consistency was high (Cronbach's alpha 0.67 - 0.91, given the age interval assessed, 82% overall, n = 1994) and test-retest reliability over 1 to 3-week intervals was 94%, n = 367. Overall concurrent validity using the Child Behavior Checklist (CBCL) and the Vineland Social-Emotional Early Childhood Scale (SEEC) was 93% (range 81% to 95%). Sensitivity was 78% (range 71% to 85%) and specificity 95% (range 90% to 98%) using individually assessed Receiver Operating Curve–based cut-off points (Pollock & Horrocks, 2008; Squires et al., 2001).

Each ASQ:SE scale has at least 21 items that can be answered by "Most of the time," "Sometimes," and "Rarely or never," which receive numerical values (0, 5, and 10). To be more comprehensive, a fourth column asks respondents to indicate whether the behavior is a concern to them. High scores indicate problem behaviors. The results can be used to identify both children with delays and as a monitoring tool for emotional and social development.

Nowadays, ASQ:SE is seen as one of the most accurate and cost-effective screening tools for use with children from birth to 5 years. The ASQ:SE is widely

used in early intervention, child care, and home visitation programs, as well as within child protection settings and by health care professionals.

3. Objectives

The present thesis had the following objectives.

- To investigate the main psychometric properties of the ASQ:SE using both Classical Test Theory (CTT) methodology and Item Response Theory (IRT).
- To present an overview of child development in daycare services using the data that were gathered from this longitudinal project.

The specific objectives were:

- To explore factorial models of the ASQ:SE and discuss its psychometric validity.
- To analyze multidimensional IRT models and compute the characteristics of each (difficulty and discrimination).
- To provide preliminary cutoff points to identify children who should be referred for further evaluation for specific developmental concerns.
- To preliminarily explore the effects of sociodemographic variables on children's socioemotional development.

4. Articles section

This thesis comprises five articles in which the following studies were conducted. Initially, based on CTT, the factor structure of the ASQ:SE was estimated using Cronbach's α, and preliminary cutoff points were computed, based on means and standard deviations. This article also discusses the specifics of emotional behaviors to differentiate "emotional behaviors" from "social behaviors"

The second article extended investigation of the factor structure of the ASQ:SE. Several models were estimated (e.g., multidimensional, bifactorial, and second-order), and some suggestions were made regarding the structure of the ASQ:SE for 5-year-old children.

In the third article, multidimensional IRT was used for two purposes: (*i*) to confirm the dimensionality of all scales of the ASQ:SE and (*ii*) to verify its item characteristics. This allowed the exploration of differences between the Graded Response Model (GRM/Samejima) and Generalized Partial Credit Model (GPCM) and discussion of the goodness-of-fit of IRT models.

In the fourth article, the focus partially shifted from psychometric properties to the preliminary results of a longitudinal evaluation of 6,530 children who were assessed over 2 consecutive years. This study allowed the determination of whether children in the public education network in Rio de Janeiro achieved some specific developmental milestones. This was very important because it, preliminarily, verified the impact of poverty on the children's development.

In the fifth article, we included the results of the ASQ:BR, with an objective that was similar to the fourth article but in which the results from 596 children were analyzed using a Bayesian approach.

Finally, despite the Brazilian tradition of dividing the psychometric field into "measurement aspects" and "evaluation aspects", the present thesis sought to treat both concepts together as a single unit.

.

ARTICLE 1

Anunciação, L., Chen, C-Y, Filgueiras, A, Squires, J., & Landeira-Fernandez, J. ASQ:SE: Adapting to a daycare context, factorial structure and preliminary standards. (Manuscript submitted for publication)

Status: The first journal has received the submission, the journal's website reported that the manuscript has been accepted but it was archived because the lacking of documents. After contacting the editor, no answer was provided. The second journal has received the submission and the manuscript is under review (second round).

Abstract

Background

Healthy emotional and social development is important in the overall development of a child. These dimensions correlate with positive outcomes later in life. However, in Brazil, there is a lack of standardized instruments to evaluate these aspects in young children.

Objective

The aims of this study were to evaluate the psychometric properties of the Brazilian Ages and Stages Social and Emotional Questionnaire (ASQ:SE) and to discuss child development based on its results.

Methods

Data were gathered in 2011- 2012 from 94,768 children, aged 0-5 years old and enrolled in public daycare centers in Brazil. An Exploratory Factor Analysis was conducted by employing Principal Component Analysis with Varimax rotation. T-tests were used to compare the mean score of boys and girls and Cohen's d was used to determine the effect size. Normative cutoffs were developed based on extreme values.

Results

Results indicated the existence of two stable dimensions and they were named "social" and "emotional". All scales had good internal consistency as measured by Cronbach's alpha higher than .70. The t-test revealed significant gender differences starting at 18 months, but the effect size was predominately small ($d < 0.3$).

Conclusion

We found that the translated and adjusted ASQ:SE was a valid and feasible procedure for the collection of data on the developmental social and emotional status in infants and young children.

Keywords: Child development; Psychometrics; Public policies; Infancy; Social-emotional screening

Introduction

Early childhood is a priority issue in social and political agenda in Brazil. This can be seen in the body of laws, public policies and economic practices that prioritizes child development (Brasil, 1988; Tatagiba, 2010). However, in Brazil, there's a lack of measurement instruments (e.g., tests and scales) to evaluate child development. This is especially valid for children enrolled in public daycare centers (Fioravanti-bastos, Filgueiras, Lucia, & Moura, 2016) and for instruments to assess social and emotional development. These two aspects are important predictors of mental health, cognitive and executive abilities associated with academic process (reading, mathematics), planning, attention flexibility and impulse control (Brooks-Gunn & Duncan, 1997; Bull, Espy, & Wiebe, 2008; Carlson, 2005; Gruendel, 2015). This statement is also important when considering children aged 0-5 years old. On this subject, it is important to develop indicators that can quantify, in a validated, objective and standardized way, social and emotional abilities of children at age (Carlson, 2005).

Among the various instruments that evaluate these mental functions, we highlight the Ages and Stages Questionnaire - Social and Emotional (ASQ:SE) (C. M. Brown, Copeland, Sucharew, & Kahn, 2012; Squires et al., 2001). ASQ:SE is a screening tool related to social and emotional aspects of child development. ASQ:SE has eight questionnaires, that is appropriate for children aged 6 months to 5.5 years old. Each questionnaire evaluates an age group and contain up to 33 Likert-type items, in which the respondent signals whether the child performs a "rarely", "sometimes" or "often" the target behavior. This checklist is coded with 0, 5 or 10 and the total score is the sum of the responses and is a measure of child development. Results above the defined cutoff point are considered "at risk", where

further evaluations are recommended (Gedler, 2004; Jee et al., 2010; Squires et al., 2001). The development of ASQ:SE was planned to address social and emotional characteristics, such as self-control, conformity, communication, adaptation, autonomy, affection and interaction. According to Squires et al. (2001), these aspects are part of social and emotional skills or competencies.

In the original study, the instrument had adequate psychometric properties (Gedler, 2004; Heo & Squires, 2012; Jee et al., 2010; Squires et al., 2001). In that study, performed between 1995 and 2001, 3,014 children were evaluated. Among the several indicators found in this study, the diagnostic capacity was calculated by the ROC curve, with a sensitivity between 0.75 and 0.89 and a general specificity of 0.92 when considering some development delay. In addition, 97% of the participants at that time evaluated the instrument as easily understood and used (Goldstein & Brooks, 2013; Squires et al., 2001). The test-retest procedure revealed a correlation of 0.94, indicating a high reliability and stability of ASQ:SE (Squires & Bricker, 2009; Squires et al., 2001).

ASQ:SE has already been adapted to Spain (Squires et al., 2001) and South Korea (Heo & Squires, 2012) and preliminary studies with children in Turkey are found in the literature (Kucuker, Kapci, & Uslu, 2011). However, ASQ:SE has not yet been translated into (Brazilian) Portuguese. Therefore, one of the objectives of the present study was to translate this instrument into Portuguese, adapting it to daycare centers context. Thus, there was a need to adapt some groups of items originally intended for parents to daycare professionals could respond.

Another objective of the present study was to verify the factorial structure of each of the scales applied in a daycare context, taking into account the

adaptations made to verify the possibility of each scale presents a two factor solution (one related to social aspects and another related to emotional aspects). Finally, the last aim was the construction of preliminary standards based on the results achieved in this study.

Method

Participants

Participants were children enrolled in public daycare centers in Rio de Janeiro city during the years 2011 and 2012. In total, 468 daycare centers, in 10 Regional Education Coordination (CRE, in Portuguese) composed the sample and fed the database with 94,768 participants. Boys were 52% of the sample, while girls were 48%.

The process of adaptation for Brazilian daycare centers

The Brazilian version of the ASQ:SE was based on the original ASQ:SE. Initially, each questionnaire was translated into Brazilian Portuguese by three independent native Portuguese speakers with professional experience in English–Portuguese translation. Each translated item was then evaluated by a multidisciplinary panel of specialists with a high level of English fluency and different expertise in psychometrics and cross-cultural adaptation instruments, public child daycare systems, child development and education, economics, and public programs for low-income families.

This multidisciplinary panel was also allowed to change any of the translated items. Conceptual equivalence, cultural adaptation, and language idiosyncrasies were considered whenever necessary. Although efforts were made

to maintain the exact meaning of each item, one item was changed to better fit the Brazilian context: “Does your baby like to play games like Peeka-boo?” (12 months) to “Cadê o queijinho que estava aqui”, once it keeps the original meaning. Because the Brazilian version of the ASQ:SE was made to be answered by daycare professionals, we added the instruction “Ask for parents in case of needing” to 5 group of items (sleeping and eating problems questions; make sure an adult is near when exploring new places, seem too friendly with strangers and stay away from dangerous things) because some children behaviors could not be properly evaluated in daycare centers. It was the case of “Does your child check to make sure you are near when exploring new places, such as a park or a friend’s home?” (18 months) and “Does your child sleep at least 10 hours in a 24-hour period?” (24 months).

At the end of this phase, a preliminary Brazilian ASQ:SE version was then back-translated into English by a native American English speaker with high fluency in Brazilian Portuguese. The back-translated (Brazilian) ASQ:SE and original ASQ:SE were examined by three native American speakers and the multidisciplinary panel. Only minimal differences were detected and changes were made when necessary.

After this process, an initial pilot test was performed with 120 children from different public child daycare centers. Caregivers were responsible for completing the questionnaire with minimal training so problems of understanding and item comprehension could be detected. Comments and suggestions from caregivers were evaluated by the multidisciplinary panel, and a few of the suggestions were incorporated into the final version of the Brazilian ASQ:SE questionnaires.

Procedures

The directors of the 468 public daycare centers were invited to participate in a 8 hours training day in previously scheduled dates defined together with the Secretary of Education of Rio de Janeiro. Each meeting had approximately 30 directors, who were trained in the application of the ASQ:SE and committed to disseminate the instructions to the professionals of each daycare center unit. Brazilian daycare centers typically feature several activity rooms, which are organized by age group. Each class has approximately 30 children and 5 teachers.

Daycare professionals (teachers) were responsible for each class and they filled ASQ:SE according to their knowledge and observations of the behavior of children. As soon as the evaluation process was carried out, they sent their answers to research team from a simple website (www.upsb.com.br/ftp/ibnec) built for this purpose. The data collection was done in 2011 and 2012.

Statistical Analysis

The analyzes were performed in successive stages. At the first moment, the database conversion was performed following the structure of "0", "5" and "10" for the original scale responses (i.e., “10” when participant checked the column of “Often or Always” for negative items “5”, when the column “Sometimes” was checked, and “0” when participant checked the column of “Rarely or Never” for positive). In accordance with the manual, reversed coding was performed for some items and items were summed up to yield a total score (Squires et al., 2001). We didn’t include in the analysis one end question asking to check if the behavior described on item promotes a concern.

Once database was quite robust (n > 10,000) and because missing values were not due at random, a conservative strategy was performed and participants

with missing data were not analyzed (Dong & Peng, 2013). Furthermore, 6-month age had only 12 participants and this group has not been used in analysis.

To investigate the construct validity, an Exploratory Factorial Analysis (EFA) of the ASQ:SE was performed in a subsample. The extraction method was based on principal components and the rotation technique was Varimax. The varimax is a type of orthogonal rotation that attempts to maximize the variance of squared loadings on a factor to reduce the cross-loadings of the variables, leading to uncorrelated simple factor structures (Castro, Baltar, Selem, Marchioni, & Fisberg, 2015) and it was chosen because this is the most frequent rotation method applied in empirical research.

The Screeplot examination and factor interpretability were employed to determine the number of factors to retain. The reliability of each questionnaire was calculated by Cronbach's Alpha. Subsequently, descriptive and inferential frequentist analysis were performed to describe and to compare the results between male and female participants. The sample was described by age-group and the percentage of participants of each sex. The mean and standard deviation of the results, and the comparison between the sexes were taken with an independent Student's T-test. Cohen-d was computed to estimate the effect size.

Exploratory cutoffs were computed for the interpretation of results based on the chosen dimensional solution and using the additive method. Two criteria were calculated: 95th percentile and 99th percentile to the results. It is recommended to use the 99th percentile, once it will minimize false positives rate.

The Statistical Package for the Social Sciences (SPSS) v.22 program was used for the analyzes. The alpha level was fixed at 0.05.

Ethical procedures

The Research Ethics Committee of the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) approved this research.

Results

Participants were 94,768 children (52% of boys and 48% of girls), distributed in seven age groups. Table 1 reports the sample characteristics.

Table 1. Descriptive statistics for the study and demographic characteristics of participants

	Participants						Mean age (months)			
	Total		Male		Female		Male		Female	
	n	%	n	%	n	%	M	SD	M	SD
12m	1,041	1.1%	543	52.2%	498	47.8	13.0	1.6	13.2	1.51
18m	5,727	6.0%	2,980	52.0%	2,74	48.0	18.1	1.6	18.2	1.63
24m	7,660	8.1%	4,020	52.5%	3,64	47.5	23.9	1.7	24.0	1.79
30m	11,19	11.8	5,939	53.1%	5,25	46.9	30.0	1.7	29.9	1.70
36m	18,90	19.9	9,867	52.2%	9,04	47.8	37.5	2.6	37.5	2.61
48m	27,90	29.4	14,62	52.4%	13,2	47.6	47.7	3.4	47.7	3.44
60m	22,33	23.6	11,29	50.6%	11,0	49.4	60.0	3.4	60.0	3.47
Tota	94,76	100	49,28	52.0%	45,5	48.0				

To determine how many factors to retain after EFA, the screeplot examination and the interpretability of the factors were performed (Figure 1). The results indicated two stable factors in common in all ASQ:SE questionnaires and they were labeled as “Social” and “Emotional”, because of its items content. The explained variance ranged between 29% (60 months) and 21% (18 months). In the four initially questionnaires, the first factor items were related to social skills (e.g., “to play to other children”), which was reversed in the last three questionnaires: the

first factor items were related to emotional skills (e.g., “can your child calm down within 15 minutes?”).

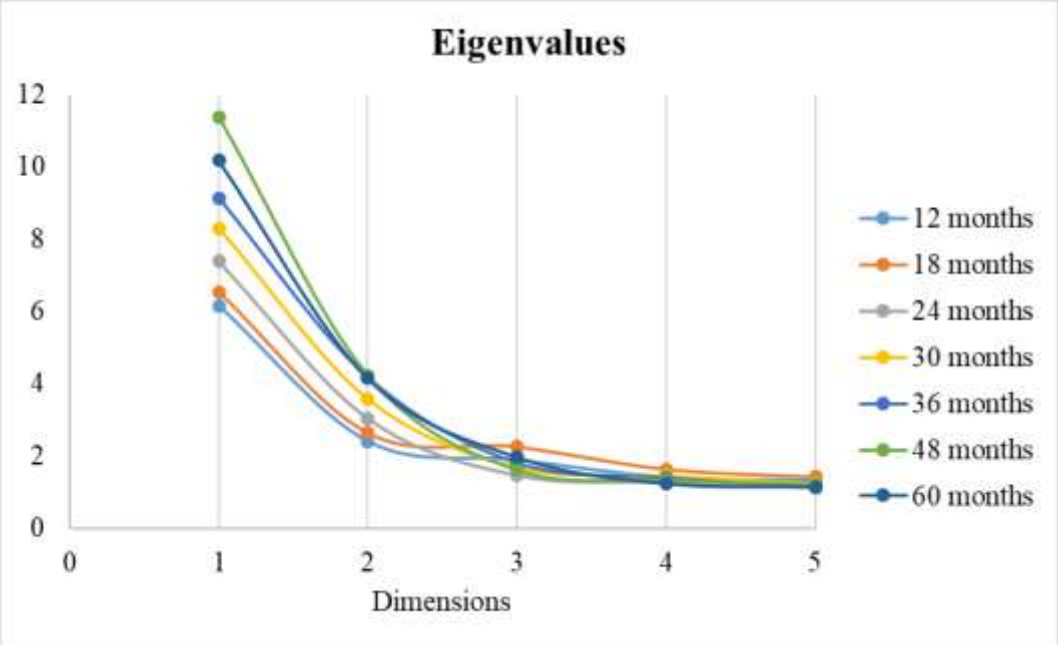

Figure 2 ASQ:SE eigenvalores

The overall Cronbach’s alpha ranged between 0.69 (12 months) and 0.85 (60 months) and its confirmed the ASQ:SE internal consistency. Table 2 presents the 12 months results as an example of all results. Other tables are included in Supplementary Information.

Table 2. Factor loadings for all items from ASQ:SE (12 months)

Ite	Content	Social	Emotional
1	Does your baby laugh or s (…)	.73	.129
20	When you talk to your bab (…)	.715	-.001
11	Is your baby interested i (…)	.611	.24
7	Does your baby like to pl (…)	.582	.296
16	Does your baby make babbl (…)	.541	.148
19	Does your baby let you kn (…)	.467	.083
3	Does your baby like to pl (…)	.411	.068
2	Does your baby look for y (…)	.333	-.01
4	Does your baby like to be (…)	.105	-.024
9	Does your baby cry, screa (…)	.204	.61
15	Does your baby have troub (…)	-.062	.592
5	When upset, can your baby (…)	.267	.505
12	Does it take longer than (…)	-.041	.494

10	Is your baby able to calm (…)	.181	.479
6	Does your baby stiffen an (…)	-.011	.421
14	Does your baby have any e (…)	-.015	.412
17	Does your baby sleep at l (…)	.096	.375
8	Is your baby's body relax (…)	.242	.368
13	Do you and your baby enjo (…)	.185	.363
18	Does your baby get consti (…)	.103	.263
21	Does your baby try to hur (…)	-.033	.251
	Explained variance	17.76%	7,74%
	Cronbach's alpha	.624	.634
	Overall Cronbach's alpha	.72	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

The table three shows the results aggregated by age group and sex. High results suggest risk in child development. Despite the statistically significant difference in most of the results, the effect size was small ($d \leq 0.3$).

Table 3. ASQ:SE results by sex

	Max	Total	Male	Female.		CI		d
		M ± SD	M ± SD	M ± SD	P-value	L	U	
12	210	24.69 ±	24.92 ±	24.44 ±	.71	-2.07	3.03	.02
18	250	26.29 ±	27.53 ±	24.95 ±	<.01	1.48	3.67	.122
24	250	24.61 ±	26.14 ±	22.92 ±	<.01	2.24	4.18	.149
30	280	29.87 ±	32.56 ±	26.83 ±	<.01	4.75	6.70	.219
36	300	29.25 ±	32.87 ±	25.30 ±	<.01	6.79	8.35	.279
48	320	28.89 ±	32.72 ±	24.68 ±	<.01	7.34	8.74	.272
60	320	39.48 ±	44.82 ±	34.01 ±	<.01	9.92	11.6	.321

Max = maximum result possible to achieve; d = Cohen's d value (effect size)

Given the difference in the indicators between boys and girls and because we work with two dimensions (social and emotional), the normative cutoff reports at Table 4 consider the age group, the sex and both of the two factors. There are two different cutoff points: the value that indicates results higher than 95% of the sample and that shows results higher than 99% of the sample. In order to avoid false positives, we recommend the use of the last one.

Table 4. Cutoff standards

	Male								Female					
	Overall		General		Social		Emot.		General		Social		Emot.	
	P. 95	P. 99	P. 95	P. 99	P. 95	P. 99	P. 95	P. 99	P. 95	P. 99	P. 95	P. 99	P. 95	P. 99
12m	65	90	65	90	30	50	45	60	65	90	35	50	40	60
18m	70	95	70	100	40	65	40	55	65	90	40	60	35	50
24m	65	100	70	100	40	65	40	60	65	95	35	60	40	60
30m	80	120	90	130	40	65	55	75	70	105	30	50	50	70
36m	85	125	90	135	35	60	70	100	75	110	30	50	55	85
48m	90	135	100	145	45	75	70	100	75	120	35	65	55	85
60m	110	150	115	160	50	75	85	115	95	135	40	65	70	100

P = Percentile

Discussion

This study investigated the psychometric properties of 7 scales of ASQ:SE using data from a large database of children enrolled in public daycare centers in Brazil. Our investigation could achieve a two factors solutions for all scales, which is theoretically convergent to ASQ:SE background. Furthermore, we investigated differences between male and female development and we created initial standards that can be used in public services.

The validity of an instrument is an index of whether (or not) it measures what it purports to measure and the question about the choice of number of how many factors must be retained is still an open question. There are, at least, 6 different suggestions in the literature, such as: 1) to consider an eigenvalue bigger

than one; 2) to explore the inflection point of the screeplot; 3) to compute the difference between each eigenvalue; 4) to verify the accumulation of the explained variance; 5) to consider the factor loading ≥ 0.30 for all items and 6) to check if each factor has, at least, 3 significant items (DeVellis, 1991; Streiner, 1994).

Once there is no official rule, it is vital to choose a factorial number that satisfies the theoretical model, without violating metric criteria (R. B. Cattell, 1966; Preacher, Zhang, Kim, & Mels, 2013). Our results gave us a strong evidence pointing to a two factors solution (Castro et al., 2015). The labels were created based on the ASQ:SE background and the items content.

The emotional factor seems has two big clusters. One of them is related to negative reactions of the children (e.g., to arc her back when picked up, cry, scream, or have tantrums for long periods of time), another related to limitations that children can show to calm down after periods of exciting activity or to move from one activity to the next.

The social factor conjugates three big item clusters. The first takes account the interest and the reaction of the child in things around her such as people or toys. The second cluster is related to the capacity of the child to communicate herself using words to describe her needs or her feelings. Finally, the third cluster involves the capacity to interact with positive affection. Items related to the capacity to follow rules or instructions or to do what an adult ask her to do showed high factor loadings in both factors.

We found statistically differences between the performance of male and female aged 18 months or older. Once the effect size was small, the difference has no practical significance (Sullivan & Feinn, 2012). However, results like these are

expected and had been well theorized and found by other researchers who used the ASQ:SE (C. M. Brown et al., 2012; Levinson, 2011). It is widely known that boys have more outsourcing behaviors, such as fighting and being more agitated than girls. This can be explained by differences in testosterone levels in boys (Lagerspetz, Björkqvist, & Peltonen, 1988), and by social stereotypes (Eccles, Wigfield, Harold, & Blumenfeld, 1993).

In relation to the normative table, the purpose of its use is to describe extreme values in general or in social or emotional dimension. Extreme results can be an important tool to identify children needing help or further evaluation. The choice of the 95th and 99th percentile is recommended in health areas when there is no gold standard measure (Apple, 2005; Coste & Pouchot, 2003; G. Singh, 2006).

Social and emotional aspects have a fundamental role in the psychological development of children. We believe that the knowledge obtained from this research could help the public management for evaluation and monitoring of services to society. This is especially important when related to childhood, given the prioritization of this stage of life in the Brazilian public agenda (Brasil, 2012; Campos, Esposito, Bhering, Gimenes, & Abuchaim, 2011; Tatagiba, 2010).

Supplementary Information

Table 5. Factor loadings for all items from ASQ:SE (18 months)

Item	Content*	Social	Emotional
19	Does your child follow s (…)	.522	.099
3	Does your child laugh or (…)	.508	.012
20	Does your child like to (…)	.476	.165
22	Does your child like to (…)	.462	.369
18	Does your child let you (…)	.460	.013
1	Does your child look at (…)	.458	.090
24	Does your child like to (…)	.453	.393
21	Does your child check to (…)	.451	-.254
6	Does your child like to (…)	.449	.115
10	Is your child interested (…)	.448	.246
4	Does your child look for (…)	.440	-.281
16	When you point at someth (…)	.411	.286
14	Do you and your child en (…)	.319	.150
5	Is your child's body rel (…)	.210	.191
9	Does your child cry, scr (…)	.063	.585
2	When you leave, does you (…)	-.025	.511
13	Does your child have tro (…)	.066	.488
23	Does your child hurt her (…)	.153	.469
7	When upset, can your chi (…)	.073	.437
8	Does your child stiffen (…)	.016	.401
11	Does your child do thing (…)	.054	.389
17	Does your child get cons (…)	.082	.356
12	Does your child have eat (…)	.115	.280
25	Does your child try to h (…)	.078	.276
15	Does your child sleep at (…)	.118	.125
	Explained variance	14.52%	6,87%
	Cronbach's alpha	.643	.582
	Overall Cronbach's alpha	.697	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

Table 6. Factor loadings for all items from ASQ:SE (14 months)

Item	Content*	Social	Emotional
3	Does your child laugh or (…)	.647	-.052
6	Does your child greet or (…)	.597	-.021
7	Does your child like to (…)	.567	.031
24	Does your child like to (…)	.564	.197
10	Is your child interested (…)	.554	.155
19	Does your child let you (…)	.546	-.042
15	When you point at someth (…)	.507	.124

1	Does your child look at (…)	.504	.046
22	Does your child like to (…)	.481	.247
18	Does your child follow s (…)	.340	.303
4	Is your child's body rel (…)	.340	.147
12	Do you and your child en (…)	.246	.240
20	Does your child check to (…)	.228	.052
11	Does your child cry, scr (…)	.099	.588
5	When you leave, does you (…)	.018	.527
21	Does your child do thing (…)	.092	.508
9	Does your child stiffen (…)	.112	.499
23	Does your child hurt him (…)	.115	.487
16	Does your child have tro (…)	.035	.467
13	Does your child have eat (…)	.083	.440
25	Does your child try to h (…)	.058	.434
2	Does your child seem too (…)	-.164	.362
8	When upset, can your chi (…)	.126	.354
17	Does your child get cons (…)	.059	.275
14	Does your child sleep at (…)	.080	.243
	Explained variance	15.54	8.04
	Cronbach's alpha	.688	.617
	Overall Cronbach's alpha	.721	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

Table 7. Factor loadings for all items from ASQ:SE (30 months)

Item	Content*	Social	Emotional
14	Is your child interested (…)	.616	.029
4	Does your child greet or (…)	.605	-.025
20	Does your child let you (…)	.598	-.055
1	Does your child look at (…)	.587	.099
18	When you point at someth (…)	.561	.062
21	Does your child follow r (…)	.540	.380
13	Does your child do what (…)	.529	.419
2	Does your child like to (…)	.518	.014
5	Does your child seem hap (…)	.518	-.021
6	Does your child like to (…)	.507	.173
12	Can your child stay with (…)	.432	.394
27	Does your child play alo (…)	.375	.094
22	Does your child check to (…)	.240	.156
17	Do you and your child en (…)	.236	.142
8	Does your child seem mor (…)	-.134	.599
9	Can your child settle hi (…)	.200	.581
28	Does your child try to h (…)	.054	.575
25	Does your child destroy (…)	.066	.563
10	Does your child cry, scr (…)	.113	.558

23	Can your child move from (…)	.342	.482
3	Does your child cling to (…)	-.169	.434
7	Does your child seem too (…)	-.148	.426
15	When upset, can your chi (…)	.193	.417
11	Does your child do thing (…)	.130	.343
24	Does your child stay awa (…)	.186	.339
26	Does your child hurt him (…)	.050	.332
16	Does your child have eat (…)	.122	.216
19	Does your child sleep at (…)	.129	.181
	Explained variance	17.89	8.42
	Cronbach's alpha	.755	.705
	Overall Cronbach's alpha	.789	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

Table 8. Factor loadings for all items from ASQ:SE (36 months)

Item	Content*	Emotional	Social
7	Can your child settle he (…)	.646	.123
29	Does your child try to h (…)	.610	.045
8	Can your child move from (…)	.608	.234
19	Does your child cry, scr (…)	.597	.082
24	Does your child destroy (…)	.591	.050
11	Does your child do what (…)	.570	.376
12	Does your child seem mor (…)	.558	-.106
5	When upset, can your chi (…)	.512	.143
18	Does your child follow r (…)	.506	.412
13	Can your child stay with (…)	.484	.338
23	Does your child stay awa (…)	.355	.114
6	Does your child seem too (…)	.332	-.126
4	Does your child cling to (…)	.331	-.149
21	Does your child do thing (…)	.313	.144
22	Does your child hurt him (…)	.310	.059
30	Does your child show an (…)	.301	-.111
20	Does your child check to (…)	.226	.113
16	Does your child sleep at (…)	.204	.078
15	Does your child have eat (…)	.199	.108
17	Does your child use word (…)	-.027	.736
26	Can your child name a fr (…)	-.016	.668
25	Does your child use word (…)	.027	.638
3	Does your child talk and (…)	-.038	.637
10	Is your child interested (…)	.082	.586
1	Does your child look at (…)	.211	.547
28	Does your child like to (…)	.118	.538
27	Do other children like t (…)	.248	.525
9	Does your child seem hap (…)	.051	.500

2	Does your child like to (…)	.078	.471
14	Do you and your child en (…)	.174	.209
	Explained variance	18.58	9.31
	Cronbach's alpha	.771	.767
	Overall Cronbach's alpha	.807	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

Table 9. Factor loadings for all items from ASQ:SE (48 months)

Item	Content*	Emotional	Social
24	Does your child follow r (…)	.713	.222
7	Can your child settle hi (…)	.656	.141
13	Does your child do what (…)	.640	.319
31	Does your child try to h (…)	.631	.075
25	Does your child destroy (…)	.598	.071
20	Can your child move from (…)	.591	.245
16	Does your child seem mor (…)	.587	-.033
8	Does your child cry, scr (…)	.583	.105
18	Can your child stay with (…)	.523	.333
4	When upset, can your chi (…)	.503	.174
6	Does your child seem too (…)	.351	-.073
32	Does your child show an (…)	.337	-.093
26	Does your child stay awa (…)	.315	.157
22	Does your child do thing (…)	.305	.176
23	Does your child hurt him (…)	.292	.037
2	Does your child cling to (…)	.290	-.053
11	Does your child have eat (…)	.201	.141
15	Does your child sleep at (…)	.168	.102
19	Does your child use word (…)	.001	.714
17	Does your child use word (…)	.003	.682
27	Can your child name a fr (…)	.009	.601
3	Does your child talk and (…)	-.077	.597
14	Does your child seem hap (…)	.079	.541
30	Does your child like to (…)	.099	.535
9	Is your child interested (…)	.076	.534
28	Does your child show con (…)	.189	.531
1	Does your child look at (…)	.246	.507
29	Do other children like t (…)	.288	.470
5	Does your child like to (…)	.087	.432
21	Does your child explore (…)	.011	.411
10	Does your child stay dry (…)	.117	.294
12	Do you and your child en (…)	.147	.240
	Explained variance	19.01	8.73

Cronbach's alpha	.807	.780
Overall Cronbach's alpha	0.839	

Item content was limited to 25 characters.

Table 10. Factor loadings for all items from ASQ:SE (60 months)

Item	Content*	Emotional	Social
24	Does your child follow r (…)	.682	.252
15	Does your child sleep at (…)	.635	.300
25	Does your child destroy (…)	.626	.087
31	Does your child try to h (…)	.616	.235
7	Can your child settle hi (…)	.612	.165
16	Does your child seem mor (…)	.596	-.130
30	Does your child like to (…)	.579	.123
9	Is your child interested (…)	.563	.075
20	Can your child move from (…)	.549	.292
13	Does your child do what (…)	.517	.331
5	Does your child like to (…)	.506	.182
22	Does your child do thing (…)	.412	.053
6	Does your child seem too (…)	.400	-.137
26	Does your child stay awa (…)	.388	.130
2	Does your child cling to (…)	.367	-.213
32	Does your child show an (…)	.357	-.101
23	Does your child hurt him (…)	.330	.065
12	Do you and your child en (…)	.286	.043
17	Does your child use word (…)	.218	.171
19	Does your child use word (…)	-.002	.679
18	Can your child stay with (…)	.039	.651
4	When upset, can your chi (…)	-.084	.596
10	Does your child stay dry (…)	.068	.563
27	Can your child name a fr (…)	.178	.551
8	Does your child cry, scr (…)	.088	.550
3	Does your child talk and (…)	-.012	.538
1	Does your child look at (…)	.233	.532
28	Does your child show con (…)	.333	.507
29	Do other children like t (…)	.126	.502
21	Does your child explore (…)	-.013	.425
14	Does your child seem hap (…)	.122	.287
11	Does your child have eat (…)	.168	.187
	Explained variance	19.91	9.11
	Cronbach's alpha	.833	.769
	Overall Cronbach's alpha	0.847	

Item content was limited to 25 characters. All Material is Copyrighted (Brookes Publishing Co)

ARTICLE 2

Anunciação, L., Pereira, D. A., Chen, C. Y, & Landeira-Fernandez, J. ASQ:SE: A factor analytic investigation of the ASQ:SE with preschool children.
(Manuscript submitted for publication. Original text was written and submitted in Portuguese and translated into English for this thesis)
Status: The journal has received the manuscript submission, which is currently undergoing a third round of review.

Abstract

Screening tools for health and education offer rapid indicators that support various practices, such as diagnostic assessment and treatment. With regard to socioemotional skills that are acquired during child development, the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) was applied to 23,334 children who were 5 years old (SD = 3 months; 50.6% male) in 2011. These children were enrolled in 625 public preschools throughout the city of Rio de Janeiro, Brazil. The goal of this study was to evaluate the validity and factor structure of the ASQ:SE. The results from the factor analysis allowed the construction of a shortened version of the tool, in which two dimensions are evaluated (emotional and social): χ^2_{433} = 5112.3, $p < .00001$, root mean square error of approximation = .049, comparative fit index = .905, Tucker Lewis index = .90.

Keywords: Child Development; Psychometrics; Public Policy.

Introduction

Aspects of social and emotional development during childhood are widely recognized as crucial for the future development of cognitive function (Briggs et al., 2012; Saarni, Campos, Camras, & Witherington, 2008). The Ages and Stages Questionnaires: Social-Emotional (ASQ:SE) is a tool that was designed to assess these skills in 6-month-old to 5-year-old children. It comprises eight questionnaires that contain Likert-type items. The sum of their items allows a rapid and objective indicator of child development with three classification possibilities: typical, questionable, and delayed (Squires, Bricker, Heo, & Twombly, 2001). In addition to the questionnaires, a manual presents its psychometric qualities, a theoretical background, and a form to apply and correct the questionnaires.

The ASQ:SE integrates the ASQ System, which includes additional tools for assessing and monitoring child development (Squires & Bricker, 2009). According to the manual, the emotional dimension refers to the ability to regulate one's own emotions in different social situations. An important feature of this dimension is the ability to recognize both one's own and others' emotions. The social dimension is related to the ability to negotiate social situations, and it is associated with a positive relationship among peers (Saarni et al., 2008).

The theoretical model has seven "subdimensions": (1) Self-regulation, (2) Conformity, (3) Communication, (4) Adapted functioning, (5) Autonomy, (6) Affection, and (7) Interaction. Self-regulation is the ability to adjust to environmental or physiological conditions. Conformity is the ability to conform with and obey orders. Communication refers to the ability to respond to verbal and non-verbal cues. Adapted functioning is the ability to cope with physiological needs. Autonomy is the ability to take initiative. Affection is the ability to show feelings and empathy for others. Interaction is the ability to respond to parents, other adults, and peers (Squires & Bricker, 2009).

Although the theoretical model is linked to child development milestones, the manual is unclear about the factor structure. The manual suggests a structure that comprises seven subdimensions, or a second-order or a bifactorial model. The present study explored the psychometric aspects of the ASQ:SE using data from 5-year-old children. This age was chosen because of an adequate sample size and because social and emotional skills are more easily discriminable in 5-year-old children than in 6-month-old infants (Gordon & Browne, 2014).

Methods

Participants

The dataset comprised 23,334 children from 625 public preschools in 10 Education Coordination Regions (Coordenadoria Regional de Educação) in the city of Rio de Janeiro, Brazil. The children were evaluated in 2011 and 2012. The sample comprised 50.6% boys and 49.4% girls. The mean age was 5 years (SD = 3 months). Each child was evaluated only once. The Research Ethics Committee of Pontifical Catholic University of Rio de Janeiro approved this study.

ASQ:SE:BR

The ASQ:SE is a series of eight questionnaires that are designed to be completed by parents or caregivers to address the emotional and social competence of young children (6, 12, 18, 24, 30, 36, 48, and 60 months old). Each scale has at least 21 items that can be answered with "Most of time," "Sometimes," and "Rarely or never," which receive numerical values. To be more comprehensive, a fourth column asks respondents to indicate whether the behavior is a concern to them. High scores indicate problem behaviors. Empirically derived cutoff scores indicate which children should be referred for further evaluation. The results can be used both to identify children with delays and as a monitoring tool for emotional and social development.

Consistent with recommendations in the literature for cross-cultural procedures, three steps were taken to translate and adapt the ASQ:SE to Brazil. The first step was completed using a parallel back-translation procedure. Each item was translated into Brazilian Portuguese by three independent native Portuguese speakers with professional experience in English-Portuguese translation. All of the translated items were evaluated by a multidisciplinary panel of specialists with a high level of English fluency and different expertise in psychometrics and cross-cultural adaptation instruments, public child daycare systems, child development and education, economics, and public programs for low-income families. This version was named the ASQ:SE(T), which was then back-translated into English by a native American English speaker who was fluent in Portuguese, thus creating the ASQ:SE:BR. In the second phase, the ASQ:SE:BR was compared with the original ASQ:SE, and minimal changes were made. The third and last step was a pilot test with 120 children from different public child daycare centers.

Procedure

A 1-day, 8-hour meeting was scheduled by the Education Secretary of the city of Rio de Janeiro, and all directors of the 468 public child daycare centers were invited to participate. In each meeting, the ASQ:SE:BR was presented by a person who was previously trained, and the directors were responsible for taking the ASQ:SE:BR back to their centers and instructing their teachers how to administer it in their classrooms. Children were evaluated by their teachers during the second semester in 2011. Each classroom had an average of 30 children.

Statistical plan

The psychometric analysis was based on four steps: (1) analysis of the theoretical models, (2) exploration of alternative models, (3) confirmation of the results obtained in the exploration, and (4) attempt to reduce the length of the questionnaire based on statistical fit.

In the first step, a Confirmatory Factor Analysis (CFA) was performed in accordance with the manual's suggestions. The CFA framework requires researchers to specify all characteristics of the hypothesized measurement model (e.g., number of factors, pattern of indicators, and factor relationships). Because of this, each item was assigned to only one factor according to the structure that is presented in Figure 1. Practical goodness-of-fit indices were used to evaluate the models, including the comparative fit index (CFI), Tucker Lewis index (TLI), and root mean square error of approximation (RMSEA). Conventional cutoffs for these indices are .9 or higher, indicating acceptable fit, and .95 or higher, indicating excellent fit. Based on the initial results, we proceeded to the second step.

An Exploratory Factor Analysis (EFA) was performed by considering one to seven dimensions. The EFA is a special case of Structural Equation Models (SEM). It has traditionally been used to explore the possible underlying factor structure of a set of observed variables without imposing a preconceived structure on the outcome (Barrett, 2007). The factor analysis rotation was conducted using the weighted least squares mean-adjusted (WLSM) in an oblique rotation.

Based on the EFA results, a new CFA was performed for the full scale and a shortened version that was developed based on the previous results. Consistent with recommendations in the literature, a random subsample that contained 25% of

the original base was used to perform the CFA, and a statistical equivalence test was conducted with this subsample to ensure its non-difference from the full dataset (Cabrera-Nguyen, 2010). All of the analyses were performed using R and MPLUS 7.4 software.

Results

Theoretical model

As previously noted, the theoretical model of the ASQ:SE considers seven correlated dimensions and also addresses the two main dimensions but the manual does not conclude in favor of a second-order model or a bifactorial mode (social and emotional; Squires et al., 2001). Because social and emotional development includes the child's experience, management of emotions and the ability to establish positive and rewarding relationships with others, this is justifiable.

l. The results of the seven-dimension model (Figure 2) resulted in a non-positive covariance matrix.

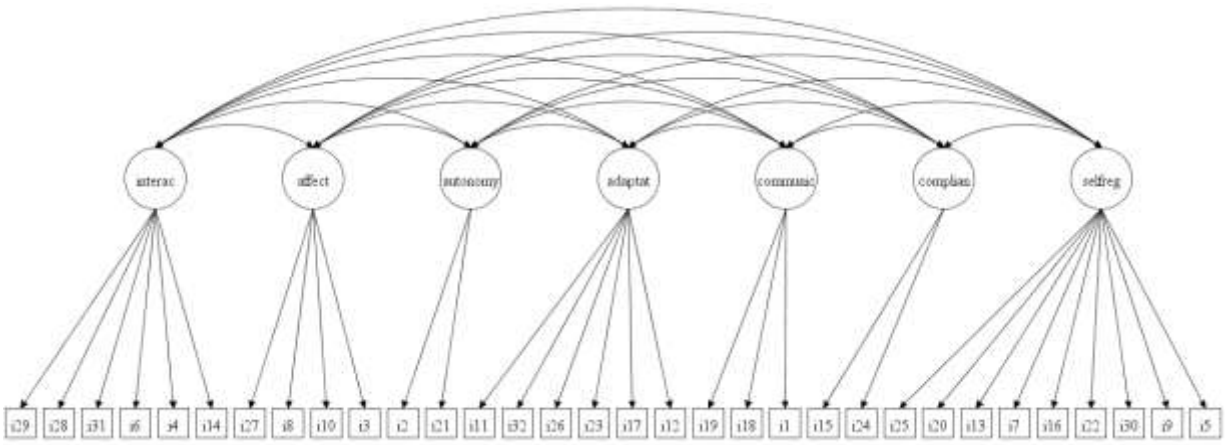

Figure 3 Oblique structure (seven dimensions).

There are some situations in which this occurs and it can be because of a perfect linear dependency of one variable on another or because of the polychoric correlations on a pairwise basis (Hair et al., 2009). The software output reported some fit results (χ^2_{443} = 45468.210, p < .00001, RMSEA = .067, CFI = .828, TLI = .807), but they were not considered. Additionally, neither the second-order (compliance, self-regulation and affect more related to emotional dimension) nor bifactorial model presented convergence (Figure 3, Figure 4).

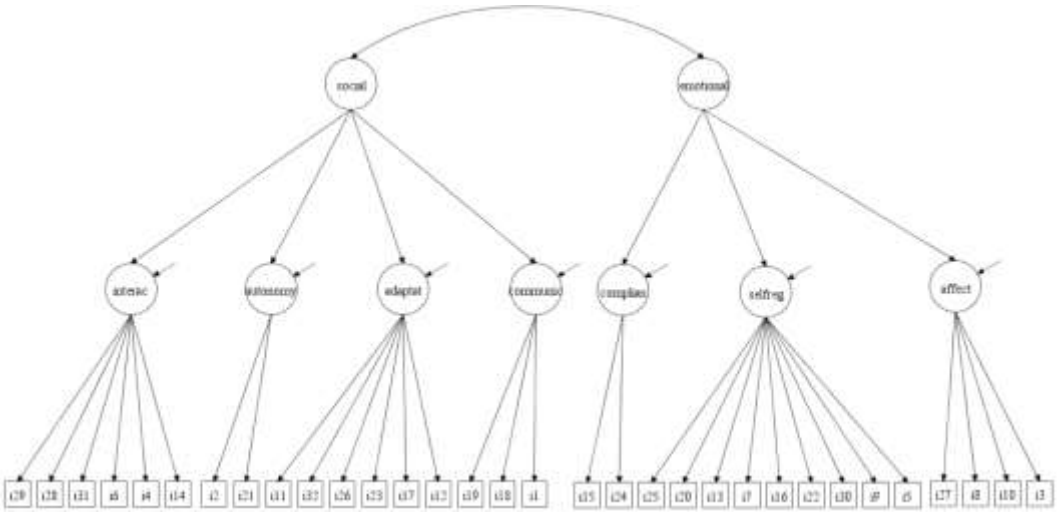

Figure 4 Hierarchical structure.

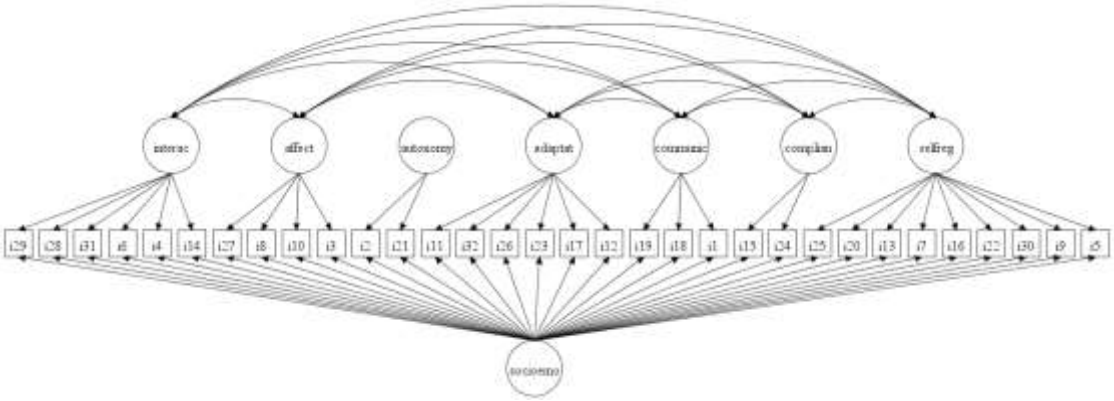

Figure 5 Bifactorial model.

Exploratory Factor Analysis

Given the results, we investigated the factor structure by imposing no restrictions on the data. The EFA is applied when one has no prior theory or measurement assumptions (Marsh, Morin, Parker, & Kaur, 2014). We checked the possibility of one to seven dimensions because the instrument was originally developed this way. Table 11 reports the results.

Table 11. EFA results (one to seven dimensions).

Model	Par	χ^2	df	P-value	CFI	TLI	RMSEA	Eigenvalue
1 Factor	32	62655.01	464	< .001	,762	.746	.077	10.3
2 Factors	63	20990.55	433	< .001	.921	.91	.046	4.1

3 Factors	93	12956.25	403	< .001	.952	.941	.037	1.9
4 Factors	122	8970.976	374	< .001	.967	.956	.032	1.2
5 Factors	150	6569.615	346	< .001	.976	.966	.028	1.2
6 Factors	177	5050.49	319	< .001	.982	.972	.026	1
7 Factors	203	3910.376	293	< .001	.986	.977	.024	.9

From two dimensions onward, the models could not be discarded by the current goodness-of-fit. The two-dimensions factors correlate to .35 and 20 items were in the first dimension and 12 in the second dimension

In the three-dimensional model, the first factor correlated with the second factor at .17 and the third factor at .36. The second factor correlated with the third factor at .45 In this model, six items were in the first, 12 were in the second, and 14 were in the third.

Table 12. Factor loadings (two to three dimensions).

	Two factors		Three factors		
Item	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
1	.481	.638	.227	.634	.494
2	.329	-.215	.672	-.119	.169
3	.172	.606	-.228	.555	.275
4	.118	.657	.013	.664	.142
5	.664	.278	.547	.324	.598
6	.407	-.108	.47	-.054	.316
7	.733	.284	.334	.248	.737
8	.335	.666	.187	.678	.338
9	.663	.155	.679	.231	.559
10	.331	.701	.262	.729	.31
11	.389	.341	.499	.411	.292
12	.418	.104	.519	.176	.316
13	.705	.459	.273	.422	.725
14	.235	.338	.076	.33	.251
15	.816	.427	.272	.367	.836
16	.596	-.072	.433	-.067	.548
17	.353	.273	.261	.29	.328
18	.31	.86	.212	.877	.308
19	.229	.797	.092	.803	.25
20	.722	.414	.335	.387	.728
21	.141	.468	.163	.504	.118

22	.586	.153	.556	.205	.504
23	.604	.206	.414	.215	.572
24	.842	.385	.224	.3	.878
25	.761	.206	.432	.189	.741
26	.496	.233	.27	.224	.489
27	.374	.624	-.005	.587	.438
28	.624	.709	.252	.69	.651
29	.425	.708	.23	.713	.433
30	.708	.239	.221	.171	.735
31	.744	.346	.257	.29	.765
32	.403	-.063	.283	-.063	.373

With regard to the theoretical background on emotional and social skills, emotional skills involve self-regulation (e.g., the child calms down alone), and social skills involve interaction and communication with other people (e.g., playing with adults). These items are also allocated in the instrument manual, which, although not confirmed in the first CFA, cannot be suppressed (Squires & Bricker, 2009). Thus, by analyzing the correlation between these items and their respective dimensions, maintaining the two-dimension factorial structure was considered more appropriate, and we labeled them as "emotional" and "social" skills.

Confirmatory Factor Analysis based on previous results

The model that was defined in the previous section was subjected to CFA, and the results were insufficient (χ^2_{463} = 9975,685, p < .00001, RMSEA = .055, CFI = .878, TLI = .869).

The χ^2 test is well known to be very sensitive to sample size (Barrett, 2007). In contrast to the continuous outcome case, the WLSM works better than the WLSM but can decrease practical fit indices, such as the CFI. However, to further explore these results, a new CFA was performed by excluding items with factor loadings < .32.

ASQ:SE (shortened version)

This shortened version was constructed by excluding item 2 ("Does your child cling to you more than you expect?"). At this point, the results met the fit

indices (χ^2_{433} = 5112.3, p < .00001, RMSEA = .049, CFI = .905, TLI = .9). The correlation between the dimensions was .52

Table 13. Results of CFA of the shortened version of the ASQ:SE (5 years old).

Item		Emotional	Social
24	Child follows rules	.84	
15	Child do what parents ask her to do	.833	
25	Child destroys or damage things on purpose	.732	
31	Child take turns and share when playing with other children	.762	
17	Child settle herself down after periods of exciting activity	.733	
20	Child move from one activity to the next with little difficulty	.776	
30	Child tries to hurt other children	.686	
13	Child stays with activities she enjoys	.754	
5	When upset, child calm down	.673	
9	Child cries	.642	
23	Child hurts herself on purpose	.58	
16	Child seems more active than other children her age	.508	
22	Child does things over and over	.559	
26	Child stays away from dangerous things	.508	
12	Child has eating problems	.401	
6	Child seems too friendly with strangers	.333	

32	Child shows interest or knowledge of sexual language	.351	
11	Child goes to the bathroom by herself	.464	
17	Child sleeps at least 8 hours in a 24-hour period	.368	
12	Child clings the parents more than they expect	.251	
18	Child uses words to tell you what she wants		.819
19	Child uses words to describe feelings		.729
28	Other children like to play with her		.91
29	Child likes to play with other children		.752
10	Child is interested in things around him (people and toys)		.709
8	Child seems happy		.671
4	Child talks or play with adults she knows well		.536
1	Child look at you when you talk to her		.771
27	Child shows concern for other people's feelings		.664
3	Child likes to be cuddled		.485
21	Child explores new places		.4
14	You and your child enjoy mealtimes together		.364

Item content was limited to its target behavior. All Material is Copyrighted (Brookes Publishing Co)

As shown in Table 13, the results of this CFA were psychometrically adequate, and the final version of the questionnaire comprised 19 items for the emotional dimension and 12 items for the social dimension.

Discussion

One way to monitor child development and prevent possible delays and difficulties is to continuously evaluate the child's progress. The use of instruments guarantees standardization, and there are clear criteria for the use of these tools, even in developing countries, such as Brazil (Fernald et al., 2009).

Any psychometric tool that is used to assess child development must be easy to apply and appropriate to the context. It must have adequate psychometric qualities, including validity and reliability, that are traditionally assessed using statistical techniques (Headrick & Sheng, 2013; Highqu & Arnold, 2007; Pasquali, 1996).

The present study performed extensive data analysis to verify that the items of the ASQ:SE measure emotional and social skills in children based on a cluster of behaviors that are related to compliance, self-regulation, interaction, and communication.

One question was suppressed: "Does your child cling to you more than you expect?" This decision was based on goodness-of-fit and did not consider the ASQ:SE as a whole system. One can argue that this decision was appropriate because it relied on psychometric indices. Nonetheless, this decision did not consider the setting where the ASQ:SE was administered (i.e., public daycare centers) and did not take into account the heterogeneity of the sample (Furr & Bacharach, 2008). It also did not consider that results can change, depending on the quality of data, selection of the participants, and statistical procedure that is used for analysis. Notably, all of the ASQ:SE items were designed to identify social and emotional difficulties, and the suppression of one item may change the way the ASQ-SE represents these skills.

Finally, we recommend using the full version of the ASQ:SE and apply the shortened version only in exceptional cases. The ASQ:SE is relatively easy to apply and provides reliable and accurate results. Such evidence-based screening tools that include parental reports can help both parents and healthcare professionals discuss a child's development in a systematic way.

ARTICLE 3

Anunciação, l., Chieh-Yu, C., Squires, J., &J. Landeira-Fernandez. Evaluation of the Brazilian Ages and Stages Questionnaires: Social-Emotional (ASQ:SE:BR) using Multidimensional Item Response Theory. (Manuscript submitted for publication)

Status: The journal has received the submission and the manuscript is under review.

Abstract

Background

Emotional and social competencies in the earliest years are vital to better health outcomes. Psychometrically adequate screening tests are needed to provide evidence for a broad range of programs. The present study investigated the psychometric properties of the Brazilian version of Ages and Stages Questionnaire: Social-Emotional (ASQ:SE:BR) using data that were gathered from public daycare centers in Brazil.

Method

A total of 54,583 children (53% boys, 12-60 months old) in 468 public daycare centers in Brazil were assessed using the ASQ:SE:BR. The Graded Response Model and the Generalized Partial Credit Model were considered for exploratory analysis. After a comparison of each model based on Akaike's Information Criterion and Schwarz's Bayesian Information Criterion, the Graded Response Model was used for confirmatory analysis.

Results

A two-dimensional solution (emotional and social) was achieved for all age ranges. The Root Mean Squared Error of Approximation (RMSEA) was less than .8, and the Standardized Root Mean Square Residual (SRMRS) was less than .1. Discrimination (slopes) and locations (thresholds) were calculated for each scale and dimension.

Conclusions

The findings provide evidence that supports the ASQ:SE:BR as a psychometrically adequate indicator of the development of emotional and social dimensions in children who are enrolled in public daycare centers in Brazil, suggesting that this instrument is a useful screening tool for this population.

Keywords: Child Development; Psychometrics; Day care; Development

Key Practitioner Message

- It is recommended that all children receive developmental screening at certain ages
- Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) was developed to measure social and emotional development in early childhood
- ASQ:SE was adapted to public Brazilian daycare centers and it was used as a screening tool for children ages 1-5 years
- A two-dimensional confirmatory item response theory model has shown an adequate fit for all age groups
- ASQ:SE assess social and emotional dimension in young children and can be used to identify children at risk of developmental delay and the need for referral for evaluation

Introduction

The importance of healthy social and emotional development in young children (0-5 years old) is well established in the literature, and these competencies are increasingly recognized as decisive for children's success (Darling-Churchill & Lippman, 2016; Schimmenti & Bifulco, 2015). Several research studies suggest that early emotional and social competencies facilitate children's successful interactions with others and are linked to later academic achievement, better global and mental health, less criminal activity, and less substance use (Halle & Darling-Churchill, 2016; Kalvin, Bierman, & Gatzke-Kopp, 2016). By contrast, inadequate levels of these competencies are recognized as contributing to public health problems, including violence and substance abuse (D. E. Jones, Greenberg, & Crowley, 2015). These findings impact policies, programs, and investments for young children in both high and low income countries (Carey, 2001; Sabanathan, Wills, & Gladstone, 2015).

As countries develop policies and make investments in early childhood programs, the dearth of empirical studies that are related to measuring social emotional competence in young children, especially in low-income settings, has become apparent. Over the last decade, conceptual frameworks to guide such measurements and the development of several social emotional screening measures for tracking at the population level have provided a strong starting point for empirical studies on a national, regional, and global scale.

Accurate and reliable measurements of early childhood social emotional competence can help (1) inform sound policy formation, (2) identify where additional investments are needed, (3) inform curricula, instruction, and teacher training, and (4) identify children who are at risk of poor school achievement at the individual or group level (Carey, 2001; Sabanathan et al., 2015). However, the development of psychometrically adequate measures is seen as a challenge for the field (Halle & Darling-Churchill, 2016). Such challenges include identifying abilities within these domains that are predominant during early and later stages of development, measuring positive development instead of only pathological behaviors, adapting measures that can be easy and quickly administered and are sensitive to cultural variations, and capturing the diversity of children's developing competencies (Darling-Churchill & Lippman, 2016; Gudmundsson, 2009).

In the field of measurement, the Ages & Stages Questionnaire: Social-Emotional (ASQ:SE) is a low-cost screening instrument that can accurately measure the emotional and social competence of young children (Squires et al., 1997). The ASQ:SE is recognized as a reliable and valid screening instrument to identify potential social emotional competence problems and delays among children who are 2 months to 5 years of age, in addition to being an easy instrument to apply that requires minimal training. Because of these characteristics, the ASQ:SE was chosen as the outcome measure in public Brazilian daycare centers as part of a 3-year public initiative to identify children with developmental delays (Filgueiras, Pires, & Landeira-Fernandez, 2014; Squires et al., 1997).

Despite solid initial psychometric studies, the properties of the ASQ:SE for this specific population have not been studied. Therefore, the aims of the present study were (1) to test the psychometric properties (validity and reliability) of the Brazilian version of the ASQ:SE (ASQ:SE:BR) by comparing two models from Item Response Theory (IRT) and (2) to verify its appropriateness with a group of typically developing children who were enrolled in public daycare centers in Brazil.

Method

Participants

Demographic details are presented in Table 14. The sample consisted of children who were enrolled in all 468 public daycare centers in the city of Rio de Janeiro, Brazil, in 2011.

Table 14. Characteristics of participants.

		Male	**Female**	**Mean age (years [SD])**	
Age	**Total**	*n (%)*	*n (%)*	*Male*	*Female*
12 months	336	184 (55%)	152 (45%)	12.08 (1.65)	12.26 (1.63)
18 months	2,611	1,383 (53%)	1,228 (47%)	18.13 (1.51)	18.1 (1.56)
24 months	3,776	2,030 (54%)	1,746 (46%)	23.67 (1.72)	23.69 (1.72)
30 months	5,064	2,647 (52%)	2,417 (48%)	29.7 (1.71)	29.68 (1.7)
36 months	7,979	4,198 (53%)	3,781 (47%)	37.24 (2.61)	37.18 (2.64)
48 months	12,473	6,495 (52%)	5,978 (48%)	47.69 (3.55)	47.76 (3.52)
60 months	22,331	11,291 (51%)	11,040 (49%)	60 (3.44)	60 (3.47)

Total	54,583	28,228 (53%)	26,345 (47%)	45.95 (14.37)	46.53 (14.27)

Ethical approval

Ethical approval for this study was granted by the Ethical Committee of Pontifícia Universidade Católica do Rio de Janeiro (Rio de Janeiro/Brazil).

ASQ:SE:BR

The ASQ:SE is a series of eight questionnaires that are designed to be completed by parents or caregivers to address the emotional and social competence of young children (6, 12, 18, 24, 30, 36, 48, and 60 months old). Each scale has at least 21 items that can be answered with "Most of time," "Sometimes," and "Rarely or never," which receive numerical values. To be more comprehensive, a fourth column asks respondents to indicate whether the behavior is a concern to them. High scores indicate problem behaviors. Empirically derived cutoff scores indicate which children should be referred for further evaluation. The results can be used both to identify children with delays and as a monitoring tool for emotional and social development.

Consistent with recommendations in the literature for cross-cultural procedures, three steps were taken to translate and adapt the ASQ:SE to Brazil. The first was completed using a parallel back-translation procedure. Each item was translated into Brazilian Portuguese by three independent native Portuguese speakers with professional experience in English-Portuguese translation. All of the translated items were evaluated by a multidisciplinary panel of specialists with a high level of English fluency and different expertise in psychometrics and cross-cultural adaptation instruments, public child daycare systems, child development and education, economics, and public programs for low-income families. This version was named the ASQ:SE(T), which was then back-translated into English by a native American English speaker who was fluent in Portuguese, thus creating the ASQ:SE:BR. In the second phase, the ASQ:SE:BR was compared to the original ASQ:SE, and minimal changes were made. The third and last phase was a pilot test with 120 children from different public child daycare centers.

Procedure

A 1-day, 8-hour meeting was scheduled by the Education Secretary of the City of Rio de Janeiro, and all directors of the 468 public child daycare centers were invited to participate. In each meeting, the ASQ:SE:BR was presented by a person who was previously trained, and the directors were responsible for taking the ASQ:SE:BR back to their centers and instructing their teachers how to administer it in their classrooms. Children were evaluated by their teachers during the second semester in 2011. Each classroom had an average of 30 children.

Statistical Analysis

The goal of the analyses was to explore and confirm the factor structure of the ASQ:SE:BR using an IRT model that was related to the nature of the data that were collected (ordinal). The Samejima's Graded Response Model (GRM) and Generalized Partial Credit Model (GPCM) were considered because they are flexible polytomous models, with fewer assumptions, which allows for separate discrimination parameters and separate category response parameters to be estimated for each item.

The GRM is an extension of the two-parameter logistic model (2PL) and is particularly suited for ordered item responses. Each item is described by a slope parameter (a) and a between category threshold/location parameter (b). Items with higher slope parameters are more informative (this measure is similar to an Item-total correlation). The ASQ:SE:BR has three categories, and each item of the ASQ:SE has one slope and two thresholds for each dimension (the latent trait level in which the probability of answering "not at all" *vs*. "sometimes" and "all the time" is equal [50/50; 1 *vs*. 2-3] and the latent trait level that is necessary to have a 50% chance of answering "all the time" instead of "sometimes" and "not at all" [1, 2 *vs*. 3].

The GPCM extends the Partial Credit Model (PCM) by introducing a variable discrimination parameter (a) for each item, similar to the GRM. Once the GPCM dichotomizes adjacent categories in thresholds ("not at all" *vs*. "sometimes"; "sometimes" *vs*. "all the time") to execute a local estimation, there is no guarantee that categories of thresholds will be ordered as it was created.

The data analysis was conducted in the following successive steps:

(1) ASQ:SE:BR scores were transformed from 0 (Always), 5 (Sometimes), and 10 (Rarely or never) to 2 (Always), 1 (Sometimes),

and 0 (Never) because IRT models follow a monotonically increasing shape.

(2) Datasets were split into two random samples: one for exploratory analysis and another for confirmatory analysis. The ratio of participants/items was 10:1.

(3) A two-factor multidimensional exploratory IRT was modeled with the GRM and GPCM.

(4) Goodness-of-fit (Cook, Kallen, & Amtmann, 2009) models were considered adequate if Root Mean Squared Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMSR) were not greater than .08 or .1, respectively (Arias, Nuñez, Martínez-Molina, Ponce, & Arias, 2016; Cai & Hansen, 2013; Elafros et al., 2015; Maydeu-Olivares, 2013).

(5) Because the GRM and GPCM are non-nested models, Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC) were used to compare both. For these analyses, smaller values indicate a better model and are ideal for non-nested models.

(6) All items of each scale were assigned to either factor based on the first exploratory results, and a two-factor confirmatory IRT factor analysis was performed with the confirmatory subset that was modeled with the GRM.

(7) The fit was checked as in item 4 above.

(8) The results were then interpreted if their discrimination parameter was $\geq .065$.

The exploratory analyses were performed using an oblimin rotation solution for better interpretation, and the Metropolis-Hastings Robbins-Monro (MH-RM) method was defined for exploratory and confirmatory analyses. This technique was chosen because the MH-RM algorithm deals with the integration problem in a different way than the traditional expectation–maximization (EM) algorithm approach and outperforms the latter technique. All of the statistical analyses were performed using R 3.3.3 and MIRT packages.

Results

The ASQ:SE has ordered polytomous categories, meaning that both the GPCM and GRM could be used to model its responses. To evaluate fit, the RMSEA and SRMSR were checked. The RMSEA statistic assesses discrepancies between the observed proportions and the probabilities that are expected under the model of interest. The SRMSR is based on discrepancies between observed and expected correlations for any pair of items. The GRM and GPCM performed similarly, and little advantage was found for the GRM across all intervals.

The comparison of each model was implemented using the AIC and BIC. These two statistics are based on the trade-off between model fit and complexity of the model. The AIC is derived from information theory and is designed to select the model that produces a probability distribution with the smallest discrepancy from the true distribution. The BIC results from a large sample asymptotic approximation to the full Bayesian model comparison. For both models, a smaller value indicates a better model fit (Table 15). In this analysis, the GRM had a better fit than the GPCM for all age intervals.

Table 15. Results of comparisons between GPCM and GRM models for each scale.

Age	Model	M2	df	*p*	RMSEA	SRMSR	AIC	BIC
12	GPCM	386.78	148	< .01	.04	.05	22167.47	22578.15
months	GRM	396.26	148	< .01	.04	.05	**22062.44**	**22473.11**
18	GPCM	2185.2	226	< .01	.04	.07	130427.9	131086.5
months	GRM	2353.7	226	< .01	.04	.06	**129905.8**	**130564.5**
24	GPCM	1776.5	226	< .01	.03	.05	170967.6	171655.0
months	GRM	2007.2	226	< .01	.03	.04	**169859.0**	**170546.4**
30	GPCM	4880.1	295	< .01	.04	.05	286092.9	286905.8
months	GRM	5210.1	295	< .01	.04	.04	**284357.4**	**285170.3**
36	GPCM	9601.7	346	< .01	.04	.05	474598.7	475532.6
months	GRM	9683.7	346	< .01	.04	.05	**472109.9**	**473043.8**
48	GPCM	16062	401	< .01	.04	.05	719053.1	720099.2
months	GRM	16921	401	< .01	.04	.06	**714271.8**	**715317.9**
	GPCM	16088	401	< .01	.04	.05	706487.4	707505.1

60 months	GRM	16948.9	401	< .01	.04	.04	**701789.7**	**702807.5**

After the exploratory analysis within an IRT framework, confirmatory analysis was performed to confirm the relationships between the items and the latent structure. The results are shown in Table 16. Deterioration of the indices was observed when comparing the exploratory analysis with the confirmatory analysis, which was expected because of the restrictions that were imposed on the model. Nevertheless, this weakening was not to such a degree that it would severely compromise acceptability of the fit.

Table 16. Confirmatory indices for each scale.

Age	**M2**	**df**	***p***	**RMSEA**	**SRMSR**
12 months	323.3756	167	< .01	.04	.07
18 months	783.5833	250	< .01	.07	.09
24 months	4363.343	250	< .01	.05	.07
30 months	850.9446	322	< .01	.06	.09
36 months	790.8339	375	< .01	.05	.09
48 months	1184.098	432	< .01	.06	.08
60 months	1193.136	431	< .01	.06	.08

The discrimination and location parameters were computed for each scale. Item discrimination reflects how well the item separates respondents with abilities below the item location from respondents with abilities above the item location. The cutoff of .65 is recommended to consider an item as moderately able to discriminate between abilities (Baker, 2001).

The average discrimination was calculated for each dimension:.73 and .61 at 12 months, .76 and .52 at 18 months, .59 and .81 at 24 months, .68 and .82 at 30 months, .95 and .75 at 36 months, .93 and .69 at 48 months, and 1.01 and .65 at 60 months. Table 17 presents the discrimination and threshold parameters for all age-interval. The intercepts of the GRM results were ordered from highest to lowest ("Always," "Sometimes," and "Rarely or never"), in which each value corresponds to the ability that is necessary to have a 50/50 chance of answering above the threshold parameters. Looking at the first row in Table 4, one can see that a child with a trait level of 5.8 (b1) has a 50% chance of responding "Rarely or never" or "Sometimes"/"Always." Similarly, a child with a trail level of 8.8 (b2) has a 50/50 chance of responding "Always."

Table 17. ASQ:SE items grouped by month.

#	Target behavior – Child	12 months				18 months				24 months				30 months				36 months				48 months				60 months			
		a1	a2	d1	d2	a1	a2	d1	d2	a1	a2	d1	d2	a1	a2	d1	d2	a1	a2	d1	d2	a1	a2	d1	d2	a1	a2	d1	d2
1	ughs	2.5	0	6.2	3.4	1.5	0	6.1	3.0	0	2.3	6.5	3.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	oks	.6	0	1.4	.2	*.5*	0	1.2	-.3	*.6*	0	2.3	.9	1.0	0	2.6	1.2	*.6*	0	3.1	1.6	1.0	0	3.2	1.8	.7	0	2.9	1.2
3	es to play w/ adults	1.8	0	6.0	4.1	1.2	0	6.2	3.9	0	1.5	2.7	.8	0	1.1	2.8	1.1	0	2.1	5.7	3.5	0	1.6	4.4	2.6	0	1.6	4.0	1.9
4	es to be picked up	.3	0	4.2	2.1	1.2	0	6.9	3.3	0	1.8	6.7	3.3	0	1.8	5.8	3.1	0	1.3	4.8	2.9	0	1.4	4.8	2.5	0	1.3	3.6	1.3
5	lms down when upset	0	2.4	5.4	3.3	0	1.5	3.6	2.0	1.2	0	3.3	2.0	1.4	0	4.2	2.6	1.6	0	4.8	2.6	1.6	0	4.3	2.4	1.2	0	4.0	1.8
6	ch his back	0	.8	2.8	1.7	0	1.3	3.7	2.6	1.6	0	3.8	2.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	es to play games	2.1	0	4.4	2.6	2.0	0	5.4	3.6	0	1.7	5.5	3.5	0	2.2	6.5	4.0	-	-	-	-	-	-	-	-	-	-	-	-
8	dy is relaxed	0	1.2	3.8	2.5	.8	0	2.9	2.2	0	1.4	4.2	2.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	ies/have tantrums	0	2.1	3.2	.9	0	2.4	4.5	2.1	2.6	0	4.8	2.3	1.9	0	4.0	1.7	1.9	0	4.0	1.7	1.7	0	4.1	1.7	1.4	0	3.3	1.6
10	lms down within time	0	1.5	3.2	1.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	interested in things	2.3	0	6.7	4.3	1.7	0	6.7	3.7	0	3.2	9.2	5.5	0	2.1	6.9	4.1	0	2.0	6.9	4.2	0	2.4	6.5	4.4	0	2.0	6.2	3.5
12	kes longer than 30 inutes to be fed	0	.9	3.0	1.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	joys the mealtimes	0	1.2	4.2	2.3	1.2	0	4.1	2.5	0	.9	3.5	1.8	0	.7	3.6	1.7	.8	0	3.5	1.9	*.5*	0	4.0	1.7	0	.7	3.1	.8
14	has eating problems	0	.7	3.6	2.3	0	.9	3.4	2.6	1.2	0	3.3	2.7	.7	0	3.3	2.7	*.6*	0	3.0	2.5	.7	0	3.7	3.0	1.0	0	3.4	2.7
15	has trouble sleeping	0	1.3	3.3	1.8	0	1.2	3.2	2.1	1.2	0	2.7	1.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	makes babbling sounds	1.1	0	2.2	.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	Sleeps at least 10 hours	0	.9	3.6	1.8	0	*.6*	3.0	1.9	.8	0	3.8	2.0	.9	0	5.0	3.0	.9	0	4.8	2.9	1.0	0	4.2	2.8	.9	0	3.9	2.2
18	has diarrhea	.5	0	3.0	.6	0	*.5*	2.6	.5	.7	0	3.0	1.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	let know when is hungry	1.0	0	3.4	1.8	.8	0	1.9	.6	0	1.3	2.6	1.2	0	1.4	3.5	1.9	0	3.6	7.3	4.8	0	2.6	6.4	4.2	0	2.9	7.4	4.8
20	turns his head when you talk	2.8	0	7.7	5.2	2.4	0	7.9	5.3	0	1.5	5.9	3.9	0	2.5	9.3	4.4	0	1.7	6.8	4.0	0	1.6	5.0	3.0	0	1.9	5.3	2.5
21	hurts other children	.2	0	3.0	1.6	0	.7	2.7	.9	.9	0	2.9	1.1	1.4	0	3.4	1.4	1.9	0	4.3	2.0	1.9	0	5.0	2.2	2.0	0	4.2	2.3
22	remains upset when you leave					0	1.5	4.1	2.6	1.7	0	4.5	3.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	does things over and over again					0	1.3	4.4	4.0	1.5	0	4.2	3.6	1.4	0	4.7	3.7	1.0	0	4.0	3.6	1.8	0	5.2	4.6	1.4	0	4.2	3.3
24	looks in the direction of what you pointing					1.6	0	7.4	3.5	0	2.0	5.9	4.0	0	2.4	8.1	4.7	-	-	-	-	-	-	-	-	-	-	-	-
25	follows instructions					1.6	0	5.2	2.7	0	1.3	3.9	2.2	0	2.6	6.8	3.2	2.0	0	5.2	3.1	2.5	0	5.5	2.0	3.5	0	6.2	1.9

26	makes sure you're near when exploring places	*.5*	0	2.2	.7	0	*.4*	2.2	.5	0	*.6*	2.2	.7	*.5*	0	2.3	.5	0	1.0	4.1	1.9	0	1.1	3.3	1.4
27	rts himself	0	1.5	4.5	3.9	1.5	0	5.5	4.5	1.7	0	4.7	4.1	2.3	0	6.4	5.8	*.6*	0	5.0	4.0	1.3	0	5.4	4.4
28	e to be around other ildren	1.4	0	4.4	2.9	0	2.6	6.2	4.4	0	1.1	4.3	2.9	-	-	-	-	-	-	-	-	-	-	-	-
29	ngs more than pected									.8	0	2.0	.6	*.6*	0	2.7	.7	*.6*	0	2.5	1.1	*.5*	0	1.9	.7
30	ems happy									0	1.7	6.2	3.7	0	1.6	6.7	3.7	0	1.8	6.2	3.5	0	1.8	5.4	2.9
31	more active than hers									1.4	0	2.2	.8	1.5	0	2.2	.8	1.2	0	2.3	1.2	1.2	0	2.0	.5
32	n settle himself down									1.9	0	4.3	1.9	2.4	0	4.9	2.3	2.8	0	5.9	2.8	1.9	0	4.5	1.6
33	ys with activities she joys									0	1.7	5.9	2.8	2.1	0	5.9	3.6	1.6	0	5.5	2.4	2.4	0	5.3	2.7
34	es what you ask to do									0	1.9	6.3	1.7	2.4	0	6.8	2.5	2.8	0	7.5	2.9	2.8	0	5.4	2.0
35	oves from one activity the next									1.8	0	5.7	2.8	2.7	0	6.7	3.6	2.4	0	5.7	3.2	1.8	0	4.2	2.0
36	ys away from ngerous things									.9	0	2.6	.9	.9	0	2.4	1.2	.8	0	2.7	1.6	1.5	0	4.0	2.1
37	mages things									1.9	0	4.0	2.3	1.6	0	4.7	2.5	2.3	0	5.5	3.5	2.4	0	5.4	3.1
38	es words to describe her feelings													0	2.2	3.2	1.6	0	3.6	6.2	3.3	0	1.8	3.9	1.9
39	can name a friend													0	4.0	8.8	6.9	0	2.2	5.9	4.7	2.1	0	4.3	.7
40	other child like to play with her													0	2.2	7.3	4.9	1.3	0	5.8	3.6	0	2.1	6.8	3.8
41	likes to play w/ other children													0	1.9	8.0	4.5	0	1.4	5.8	3.7	0	2.6	7.5	4.5
42	show an interest in sexual language													*.5*	0	3.7	2.2	.8	0	4.0	2.3	.9	0	3.7	2.1
43	stays dry during the day																	0	1.2	5.8	3.6	1.1	0	4.5	3.6
44	shows concern for other people's feeling																	0	1.5	3.8	1.3	0	1.3	3.2	.7

a1/a2 = Discrimination / factor loadings for each dimension; d1/d2 = Location/Difficulty.

The difference between discrimination parameters can be seen in a graphical interface, and it helps to compare this property of items. Figure 1 plots the most and least discriminative items of all of the scales (item 26, 36 months, “Can your child name a friend?”; item 4, 12 months, “Does your baby like to be picked up and held?”).

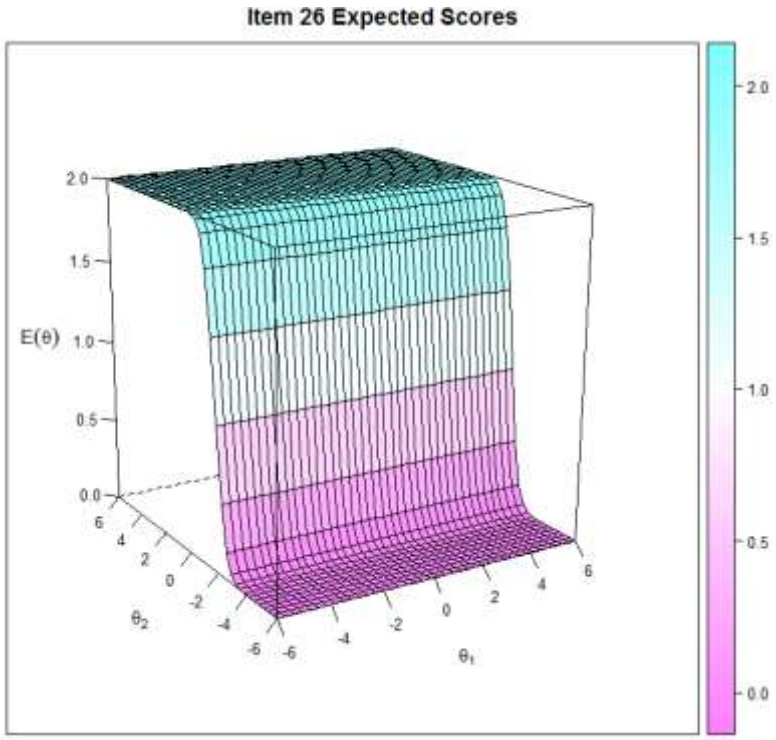

Figure 6 Item characteristics curve (1).

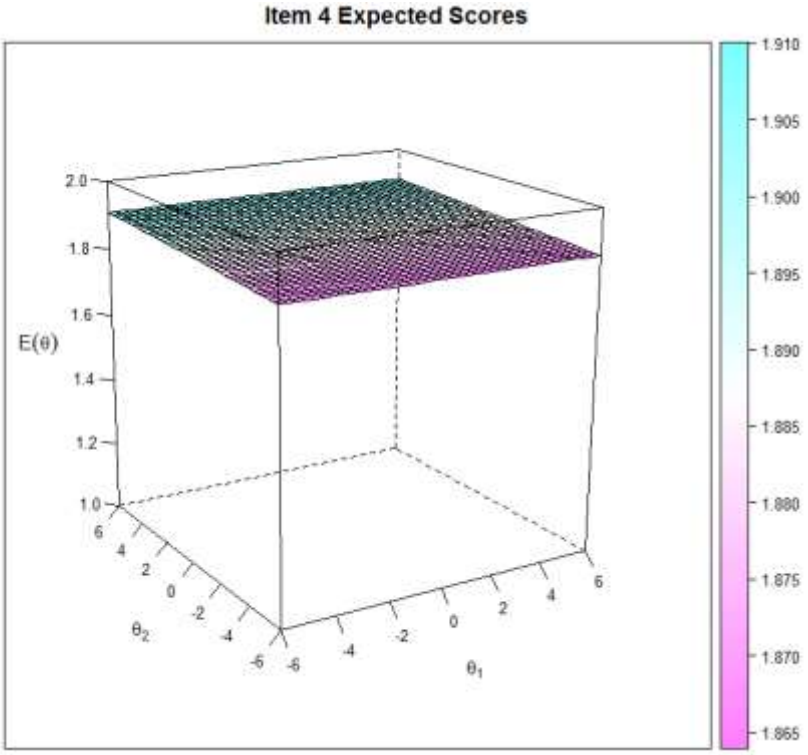

Figure 7 Item characteristics curve (2).

Finally, once the differentiation of emotional and social dimensions in development is possible, the analysis of each item has the potential to indicate the minimum ability that children must have in each age interval to have a 50% chance of overcoming the first category ("Rarely"). This analysis suggested the item, "When you talk to your baby, does he turn his head, look, or smile?" This item had the highest first threshold for children at 12 and 18 months of age. This means that children must have a minimum ability of 5.8 to have a 50/50 chance of replying to this item with "sometimes" at 12 months.

At the 24-month interval, "Does your child hurt herself on purpose?" had the highest value for the first threshold. At the 30-month interval, the higher first threshold was for "When you point at something, does your child look in the direction you are pointing?" At the 48-month interval, the higher first threshold was for "Can your child name a friend?" At the 60-month interval, the higher first threshold was for "Does your child let you know when she is hungry, hurt, or tired?"

With regard to item discrimination, children with ability to the left of the item location had a lower probability of responding correctly or endorsing the item. The item "Does your baby cry, scream, or have tantrums for long periods of time?"

was the most discriminative for the emotional dimension at 12 and 30 months. “When upset, can your baby calm down within a half hour?” was the most discriminative at 18 months. “Does your child hurt herself on purpose?” was the most discriminative at 24 months. “Can your child move from one activity to the next with little difficulty, such as from playtime to mealtime?” was most discriminative at 36 months. “Can your child settle himself down after periods of exciting activity?” and “Does your child follow simple directions?” were the most discriminative at 48 and 60 months, respectively. For the social dimension, the most discriminative items were “When you talk to your baby, does he turn his head, look, or smile?” (12 months), “When you talk to your baby, does he turn his head, look, or smile?” (24 months), “Is your baby interested in things around her, such as people, toys, and foods?” (30 months), “Can your child name a friend?” (36 months), “Does your child use words to describe her feelings and the feelings of others, such as, “I’m happy,” “I don’t like that,” or “She’s sad?” (48 months), and “Does your baby let you know when she is hungry, hurt, or tired?” (60 months).

Discussion

Latent variables are non-observable quantities that are inferred by observation of the response to items in scales or tests (Baker, 2001; Reise & Revicki, 2015). The use of IRT provides an estimate of the ability by considering both participant responses and the properties of the items, and it is used in educational and psychological research to study latent variable constructs other than ability (e.g., depression, personality, and motivation). The primary goal of this research was to explore and confirm the dimensionality of an adapted version of the ASQ:SE for Brazilian 1- to 5-year-old children who were enrolled in daycare centers. To achieve this goal, a modern approach that was based on multidimensional IRT was used. The findings provided support for two conclusions. First, a two-factor model that was based on the GRM was sufficiently valid. Second, despite the inherent relationship between emotional and social behaviors, some of them could be seen as partially independent.

Regarding the first conclusion, the fit indices in the exploratory analysis that were derived through the GRM performed better than the GPCM. Importantly, both models tended to produce similar results, but once the GRM forced the categories’

boundary parameters to be ordered (not a GPCM assumption), the results were more closely related to the theoretical background of the ASQ:SE:BR. The confirmatory analysis evaluated the degree to which the measurement hypotheses were consistent with child ASQ:SE:BR scores and were based on the error of the measurement (Arias et al., 2016; Cai & Hansen, 2013; Caprara & Cervone, 2000; Elafros et al., 2015; Maydeu-Olivares, 2013). The fit statistics supported the hypothesized two-factor structure (emotional and social). Moreover, because no model is without error because neither a theory nor a model can be a perfect mirror of reality, confirmatory analysis must combine theory and data to provide meaningful and useful results (T. Brown, 2015; Furr & Bacharach, 2008). Regarding the second conclusion, these findings indicate the possibility to view specificities in emotional and social domains. The interaction between these two competencies is well known in the literature. However, the degree to which children are effective in their social interactions with others (e.g., able to demonstrate cooperative skills and flexibility) is related to their social abilities and ability to understand the emotions of the self and others, read and interpret emotional cues, and react to others' emotions (Halle & Darling-Churchill, 2016).

Several studies have supported a link between heightened emotionality and both externalizing and internalizing behavior problems (Kalvin et al., 2016). Violent behaviors (e.g., hurting other children, adults, or animals) were related to emotional competency on the ASQ:SE:BR. By contrast, communication was assigned to social skills and was related to social ability (Schimmenti & Bifulco, 2015).

Our results suggest that the ASQ:SE has a sufficient number of items at the lower end of the scale to avoid having some children fail all items. The ASQ-SE is relatively easy to adapt to various cultures and is easy and inexpensive to use in low-resource settings. Finally, based on our experience, this instrument did not require extensive training to administer the questionnaires or implement a train-the-trainer model, in which the preschool directors received hours of training and were then able to train the teachers at their own centers without difficulty.

This research has some important limitations. This study was based on a cross-sectional measure, and the sample predominately comprised low-income children and families who attended these state-sponsored daycare centers, free of

charge. Because of this, our findings may not be generalizable to the general Brazilian population.

Overall, our data provide additional evidence of the validity and utility of the ASQ:SE. Our findings may contribute to a more accurate picture of young children's development for preschool teachers and others who are interested in preventing delays and disabilities and improving outcomes for young children. These findings may assist evaluators with identifying more valid and reliable practices for measuring children's social and emotional competence.

Acknowledgements

This project was by funded by the Brazilian National Council for Scientific and Technological Development (CNPq). All of the authors contributed to the design, data collection, analysis, and writing of this report. Jane Squires declares a conflict of interest as an author of the ASQ.

ARTICLE 4

Anunciacao, L., Squires, J., & Landeira-Fernandez, J. (2018). A Longitudinal Study of Child Development in Children Enrolled in Brazilian Public Daycare Centers. *Global Journal of Educational Studies*, *4*(1), 31. https://doi.org/10.5296/gjes.v4i1.12555

Status: Published.

Abstract

A longitudinal research study was conducted that examined aspects of child development in children who were enrolled in public daycare centers in Brazil. The participants were 596 children (1-3 years old, *n* = 51; 2-4 years old, *n* = 545) who were enrolled in 198 public daycare centers in Rio de Janeiro, Brazil. Communication, gross motor, fine motor, problem-solving, and personal-social domains were assessed using the Brazilian version of the Ages and Stages Questionnaire, 3rd edition, adapted for public child daycare centers. A Bayesian robust regression model was performed to check for gender and age differences and interactions. The findings indicated main effects of gender and age. Females had higher scores than males in the communication and personal-social domains. No interactions were found, suggesting the absence of moderation effects between age and gender. This study provides information about child development, especially in children who are enrolled in public services in Brazil.

Keywords: Child development, Longitudinal Research, Bayesian Statistics,

1. Introduction

Child development is a dynamic process and refers to the sequence of physical, language, thought, and emotional changes that occur in a child from birth to the beginning of adulthood. Monitoring developmental processes enables the determination of child and family vulnerabilities and strengths over time, which is important for supporting children with reaching their developmental potential or providing specialized services when needed. Studies have shown that the monitoring process is even more important in low-income countries because caregivers and healthcare providers may not be sufficiently equipped with knowledge about early childhood development and because low-income environments increase the odds of children having developmental delays or other early adversity (Ertem et al., 2008; Schonhaut, Armijo, Schonstedt, Alvarez, & Cordero, 2013; van Heerden, Sen, Desmond, Louw, & Richter, 2017).

In Brazil, the education system is composed of both public and private sectors. The former does not require direct payment and is funded entirely by the state. The latter requires direct payment, which is often provided by parents. Both systems have the same flow: (*i*) daycare services for infants aged 6 months to 4 years, (*ii*) preschool services between 4 and 6 years old, (*iii*) fundamental instruction between 6 and 14 years old, and (*iv*) high school between 15 and 17 years old. Rio de Janeiro's municipal public education network is also composed of Early Childhood Development Spaces, where nursery and preschool facilities are together in a space that is equipped with materials to provide stimulation for child development (Rio de Janeiro City, 2017).

Despite this overall structure, the public system faces several challenges. The rates of violation of the rights of children are still high, and there is a lack of providing for basic needs. Some children and families who access the network are residents of slums in high-poverty areas, often controlled by drug dealers and other criminal elements. This socioeconomic status places them in a situation of extreme vulnerability. Food insecurity and school dropout rates are high. This scenario impacts the entire system, and the overall quality of these daycare centers is probably not very high relative to similar centers in developed countries (Elwick, 2018; Fonseca, Sena, dos Santos, Dias, & de Melo Costa, 2013).

During 2008 and 2012, the city of Rio de Janeiro, in cooperation with the Strategic Affairs Secretariat of the Presidency of the Republic (SAE/PR),

implemented a program to reform basic education schools. One of the key aspects of this program relied on measuring child development using the Ages and Stages Questionnaire, 3rd edition (ASQ-3; (Chen et al., 2017). However, few results are reported in the literature. The aim of the present study was to investigate the development of children who were enrolled in public daycare centers in Brazil and identify possible gender and age differences and interactions in the five domains of child development, assessed by the ASQ-3.

2. Method

2.1. Participants

The sample was recruited as part of a larger educational program in Rio de Janeiro, Brazil. For this research, based on the children's age, two groups were formed. In the first group (n = 51, 53% female), the children were 1 year old at the first measurement time point (2010), and they were 3 years old at the last measurement time point (2012). The second group (n = 545, 44% female) comprised children who were 2 years old at the first measurement time point. Table 18 reports the sociodemographic characteristics of the participants.

Table 18. Sociodemographic characteristics of the participants.

	1-3 years old		2-4 years old	
Sex	No.	Proportion	No.	Proportion
Male	24	47%	304	56%
Female	27	53%	241	44%
Total	51	100%	545	100%
No. of children in each Regional Education Coordination[1]				
1	3	6%	75	14%
2	5	10%	71	13%
3	9	18%	57	10%
4	12	24%	69	13%
5	3	6%	24	4%
6	3	6%	38	7%
7	6	12%	29	5%
8	2	4%	71	13%
9	0	0%	49	9%
10	8	16%	62	11%
Total	51	100%	545	100%

[1]In 2012, Rio de Janeiro city had 10 Regional Education Coordination in all city extensions.

2.2. Measures: ASQ-BR

The ASQ is a caregiver report screening questionnaire for children aged 4 months to 5 years, which is completed with the caregiver by a trained assessor. The application time is 10-15 minutes. This instrument was designed to follow a child's development over time and provide a dependable and cost-effective strategy for the early identification of children who may require further assessment and who may benefit from some type of intervention (Bricker, Macy, Squires, & Marks, 2013).

The ASQ-3 comprises five subscales, including communication, gross motor, fine motor, problem-solving, and personal-social domains. Each domain consists of six questions, and caregivers must respond with "yes," "sometimes," or "not yet." These responses are then translated to numerical values (10, 5, and 0, respectively).

The initial work on the ASQ-3 began in 1979, and the scales have been successfully adapted to several languages, including Spanish (Armijo, Schonhaut, & Cordero, 2015), Chinese (Wei et al., 2015), Korean (Heo & Squires, 2012), and Brazilian Portuguese (Filgueiras, Pires, Maissonette, & Landeira-Fernandez, 2013). The Brazilian version of the ASQ-3 (ASQ-BR) was adapted for public child daycare centers. Psychometric studies concluded that the ASQ-BR is a valid and reliable screening tool for children who are enrolled in public daycare centers (Filgueiras et al., 2013). Because of the longitudinal follow-up in the present study, we used only selected months from the ASQ-BR (i.e., 12, 24, 26, and 48).

2.3. Procedure

A 1-day, 8-hour meeting was scheduled by the Education Secretary of the City of Rio de Janeiro, and all of the directors of the 468 public child daycare centers were invited to participate. In each meeting, the ASQ-BR was presented by a person who was previously trained, and the directors were responsible for taking the ASQ-BR back to their daycare centers and instructing their teachers on how to administer it in their classrooms. Children were evaluated by their teachers during the second semester in 2010, 2011, and 2012. Each classroom had an average of 30 children.

2.4. Statistical analysis

2.4.1. Dataset and descriptive statistics

The dataset had no missing values. Each total score was computed as the sum of all items for each dimension. This method relies on the ASQ manual. Descriptive statistics (mean, standard deviation, median, and median absolute difference) were computed for all ASQ-BR intervals. These results are reported by age and gender. A Bayesian regression model was performed to check for gender and age differences and interactions. To deal with outliers, the model error term was assumed to be Student's *T*-distributed with ν using a wide Gamma prior, as proposed in the literature (Bürkner, 2017; Ding, 2014). Males were assigned to the reference groups as well the first age of the group (i.e., for children aged 2-4, 4 was the reference category). All of the analyses were performed using *R* software (R Development Core Team, 2016) with the *brms* package (Bürkner, 2017).

2.4.2. Bayesian robust regression

A simple linear model ($y_i = \alpha + \beta_1 X_i + \epsilon_i$) can be written in terms of a probabilistic model: $\mu_i = \alpha + \beta_1 X_i$. This model defines $y_i \sim N(\mu_i, \sigma^2)$, which can be written as $y_i \sim N(\alpha + \beta_1 X_i, \sigma^2)$, where $\mu \in (-\infty, +\infty)$ is a location parameter, and $\sigma^2 > 0$ is a scale parameter. In other words, it is assumed that the dependent variable (Y) follows a normal distribution that is parametrized by a mean (μ_i) that is a linear function of X, parametrized by α, β, and the standard deviation σ (Gelman, Carlin, Stern, & Rubin, 2004; Kruschke, 2011).

Three techniques for estimating parameters from observed data are routinely performed: Ordinary Least Squares (OLS), Maximum Likelihood Estimator (MLE), and Bayesian Estimator. OLS aims to find the set of β that will minimize the squared errors, regardless of the form of the distribution of errors. If the form of the distribution of errors is known, then the MLE can be used to estimate regression coefficients by maximizing the likelihood of the data/joint probability of the observations. Finally, a Bayesian estimate allows setting a prior distribution on the parameters and using Bayes theorem to obtain a posterior distribution that is proportional to the prior and the likelihood. In this approach, instead of maximizing the likelihood function alone, it is necessary to assume a prior distribution for the parameters and use Bayes theorem (Myung, 2003; Rouder, Morey, Speckman, & Province, 2012):

$$posterior \propto likelihood * prior$$

In Bayesian inference, the likelihood function assumes a prior distribution for the estimated parameters as the following:

$$f(\alpha, \beta, \sigma^2 | Y, X) \propto \prod_{i=1}^{n} N(y_i | \alpha + \beta X_i, \sigma^2) * f_\alpha(\alpha) f_\beta(\beta) f_{\sigma^2}(\sigma^2) \quad 1$$

In psychological research, the data frequently do not meet the normality criteria because of atypical observations, which sometimes imposes different statistical procedures that seek to avoid type I and II errors. Studies have shown that statistical regression models with *T*-distributed errors produce more stable and precise results because the *T* distribution provides heavy-tails compared with the

normal. To model a *T* distribution, an additional parameter that represents the degrees of freedom (ν) is necessary (Ding, 2014; Gelman et al., 2004). It is possible to state that the error term follows a bivariate *T* distribution, $\epsilon_i \sim t_2(0_2, \Omega, \nu)$, with the density function as the following:

$$f(t; \mu, \Omega, \nu) = (2\pi)^{-1}|\Omega|^{-\frac{1}{2}}\{1 + \nu^{-1}(t-\mu)^T\Omega^{-1}(t-\mu)\}^{-\frac{\nu+2}{2}} \quad (2)$$

Because of the lack of prior knowledge, a weakly informative prior (wide Gamma) was considered for ν, whereas the Normal was considered for the regression coefficients and intercept. The No-U-Turn Sampler (NUTS) was chosen as the sampler, and a draw of 4,000 posterior samples was computed, from which the High Density Interval was computed.

2.5. Ethical approval

Ethical approval for this study was granted by the Ethical Committee of *Pontifícia Universidade Católica do Rio de Janeiro* (Rio de Janeiro, Brazil).

Results

All of the scores for boys and girls were computed separately by age according to the ASQ-BR dimensions. Table 19 presents the descriptive results.

Table 19. Longitudinal results of male and female children aged 1-3 years and 2-4 years by ASQ-3 dimension.

	Male			**Female**			**Male**			**Female**		
	1 year	2 year	3 year	1 year	2 year	3 year	2 year	3 year	4 year	2 year	3 year	4 year
Communication												
Mean	33.33	51.04	48.54	39.63	50.74	54.44	40.81	51.15	53.52	45.15	51.62	54.23
SD	15.99	13.67	10.05	15.87	15.61	5.94	17.30	9.24	10.04	17.11	8.40	9.97
Median	32.50	55.00	50.00	45.00	60.00	55.00	45.00	50.00	60.00	50.00	50.00	60.00
MAD	22.24	7.41	7.41	14.83	.00	7.41	22.24	7.41	.00	14.83	7.41	.00
Gross motor												

Mean	43.54	56.46	58.13	46.67	56.11	58.52	51.56	57.78	57.63	51.08	56.37	57.61
SD	17.41	5.61	4.12	13.52	7.38	4.56	10.54	5.68	6.47	9.88	7.79	6.49
Median	42.50	60.00	60.00	50.00	60.00	60.00	55.00	60.00	60.00	50.00	60.00	60.00
MAD	25.95	.00	.00	14.83	.00	.00	7.41	.00	.00	14.83	.00	.00
Fine motor												
Mean	37.50	47.71	48.54	42.22	52.41	53.15	43.16	50.26	46.81	44.52	50.81	50.54
SD	15.25	10.63	14.71	13.82	9.44	11.61	11.68	13.32	13.91	12.24	12.76	12.42
Median	40.00	50.00	52.50	45.00	55.00	60.00	45.00	55.00	50.00	50.00	55.00	55.00
MAD	22.24	7.41	11.12	7.41	7.41	.00	7.41	7.41	14.83	14.83	7.41	7.41
Problem solving												
Mean	30.00	46.25	51.88	36.11	50.93	54.44	40.02	52.90	50.99	42.03	52.74	51.99
SD	17.19	12.00	12.41	14.43	9.41	6.70	12.72	9.93	10.60	12.39	9.66	9.90
Median	30.00	45.00	60.00	40.00	50.00	55.00	40.00	60.00	55.00	45.00	60.00	55.00
MAD	18.53	14.83	.00	14.83	7.41	7.41	14.83	.00	7.41	7.41	.00	7.41
Personal-social												
Mean	31.46	44.17	52.71	32.41	47.78	56.85	32.92	52.01	54.67	37.95	54.17	55.62
SD	15.21	14.57	9.67	17.12	10.13	4.83	12.78	9.19	7.92	10.98	7.47	7.50
Median	30.00	47.50	55.00	30.00	50.00	60.00	35.00	55.00	60.00	40.00	55.00	60.00
MAD	14.83	11.12	7.41	22.24	7.41	.00	14.83	7.41	.00	14.83	7.41	.00

To check for differences and interactions across age and sex, a robust Bayesian regression was performed. Instead of using a normal distribution for the error term, Student's *T* distribution was used because of its fatter tails. This procedure allows outliers to have a smaller Mean Square Error in the likelihood and thus less influence on the regression results.

NUTS, a Markov chain Monte Carlo (MCMC) method, generated a large and representative sample from the posterior distribution, and the Highest Density Interval (HDI) was computed from these results. When the HDI does not include a null value, it is possible to conclude that the probability of observing such a value is less than 95% (Kruschke, 2011). This procedure is somewhat analogous to the Null Hypothesis Significance Testing confidence interval, but it allows the assumption of a probability statement of the parameter given the data that are analyzed. See Table 20 for the results.

Table 20. Longitudinal growth curve (1-3 years and 2-4 years).

	1-3 years					2-4 years				
	β	SE	l-CI	u-CI	Eff. Sample	β	SE	l-CI	u-CI	Eff. Sample
Communication										
Intercept	36.76	2.59	31.71	41.83	4311	42.14	0.66	40.86	43.46	5382
Female	4.11	3.57	-2.76	11.09	3988	3.65	1	1.72	5.62	4844
Age	7.58	2.00	3.73	11.51	4096	6.36	0.51	5.36	7.35	4869
Female:Age	-.17	2.74	-5.62	5.13	3663	-1.81	.78	-3.35	-0.28	4241
Gross motor										
Intercept	45.43	1.89	41.75	49.13	4848	52.63	.43	51.78	53.48	4498
Female	2.4	2.59	-2.74	7.36	4783	-.87	.64	-2.1	.37	3786
Age	7.29	1.46	4.46	10.12	4593	3.03	.33	2.39	3.7	4199
Female:Age	-1.37	2.01	-5.29	2.58	4329	.24	.5	-.71	1.2	3479
Fine motor										
Intercept	39.08	2.39	34.47	43.74	4394	44.94	0.69	43.6	46.28	4737
Female	4.76	3.28	-1.73	11.1	3809	.67	1.04	-1.38	2.72	3921
Age	5.53	1.85	1.89	9.16	3923	1.81	.53	.79	2.87	4327
Female:Age	-.1	2.55	-4.98	4.97	3417	1.21	.81	-.38	2.78	3625
Problem solving										
Intercept	31.83	2.33	27.19	36.41	4426	42.49	.6	41.33	43.67	5061
Female	6.19	3.25	-.16	12.52	3914	1.45	.88	-.27	3.18	4465
Age	10.93	1.81	7.44	14.52	4264	5.48	.46	4.57	6.37	4832
Female:Age	-1.75	2.51	-6.81	3.11	3654	-.5	.69	-1.88	.85	4011
Personal-social										
Intercept	32.19	2.36	27.65	36.87	4381	35.66	.55	34.61	36.73	4799
Female	1.26	3.25	-5.18	7.67	3844	4.74	.82	3.16	6.32	3579
Age	10.6	1.81	7.06	14.09	4008	10.87	.43	10.04	11.72	3688
Female:Age	1.63	2.5	-3.28	6.57	3407	-2.03	.64	-3.29	-0.79	3091

Significant main effects of sex and age were found. Females had higher scores than males in the communication and personal-social domains. The main effect of age showed that ASQ scores became higher as the child got older. We emphasize that the ASQ was developed as a screening tool and not as a longitudinal measurement. However, higher scores indicate that additional developmental milestones are attained, but the sample of behavior is small (i.e., six items per domain). No interactions were found, suggesting the absence of moderation effects between the children's age and gender. Figure 1 shows the children's growth curves.

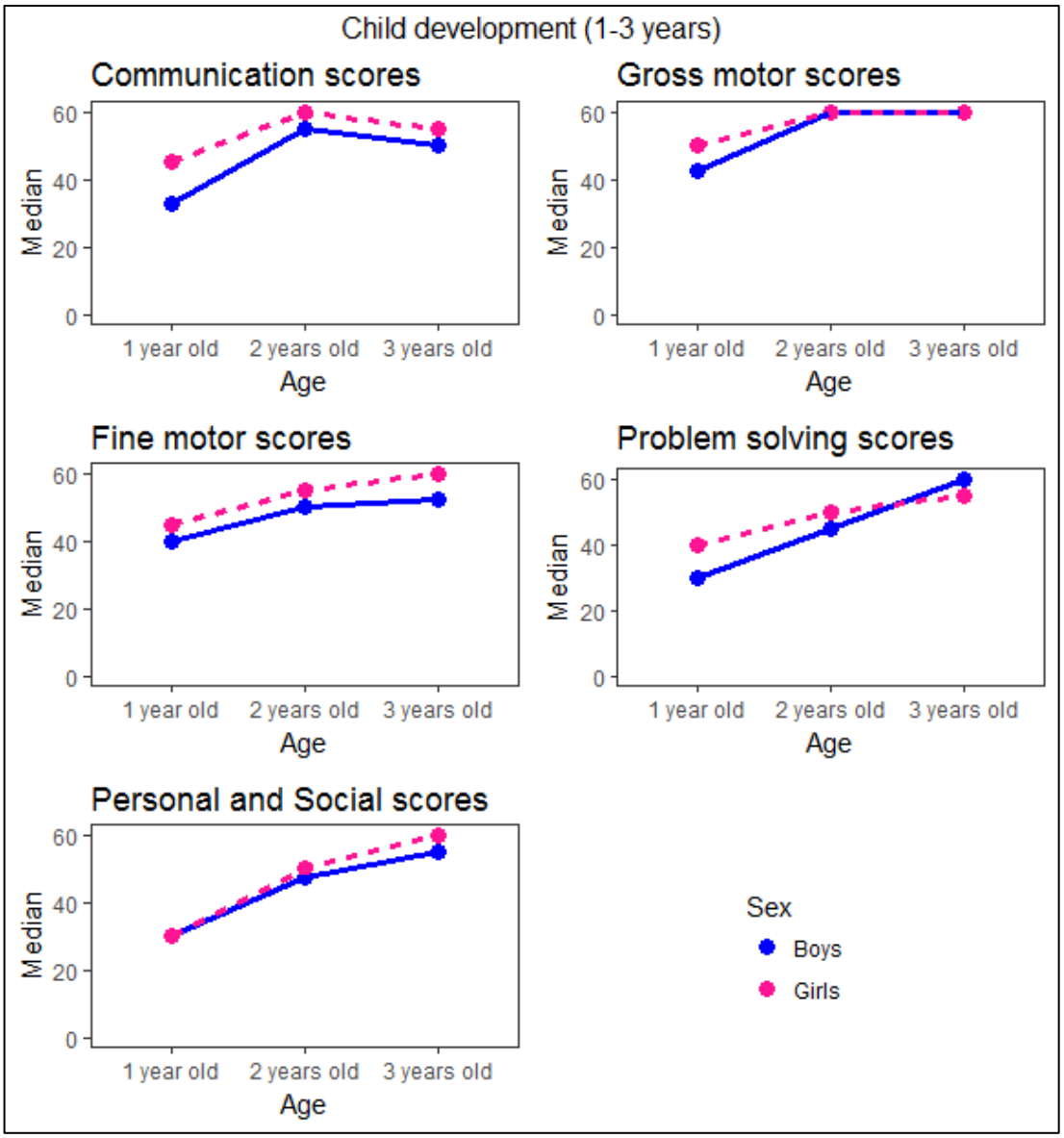

Figure 8 Children´s growth curves (1-3 years)

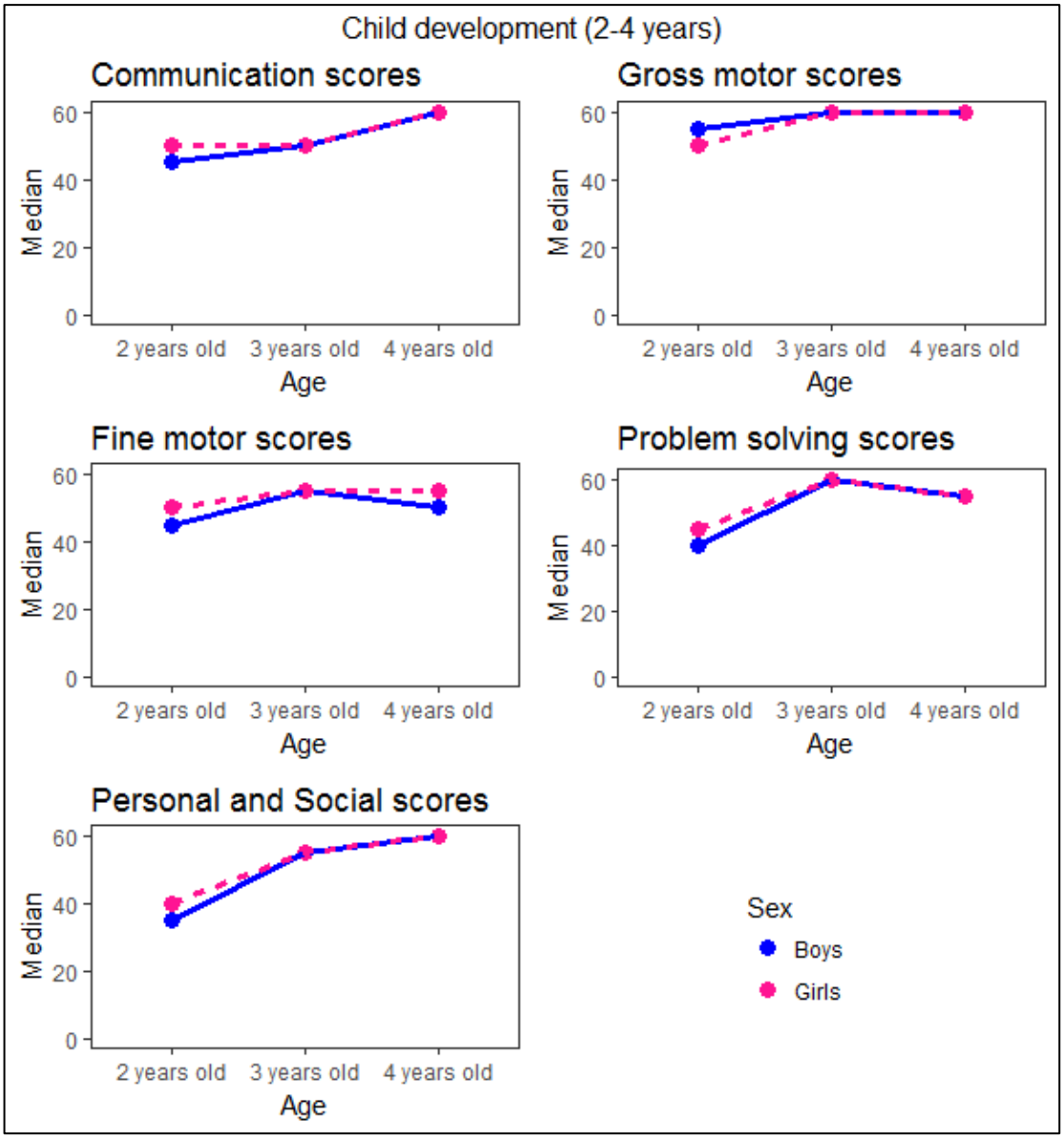

Figure 9 Children´s growth development (2-4 years)

Discussion

The study of child development in low- and middle-income countries is an important issue in the political realm and has received substantial attention in Brazil (Jannuzzi, 2002). There are at least two reasons for this emphasis. First, strong causal evidence suggests that disrupted or delayed early development has long-term negative consequences for children, which is also true for children who are exposed to multiple risks (e.g., poverty and birth trauma; (van Heerden et al., 2017; Wallander et al., 2014). Second, economic investment in children could yield high rates of return (up to nine times) in poor countries, assessed by such outcomes as income and employment rates (Campbell et al., 2014; Stenberg et al., 2014).

The present study investigated five domains of child development in children who were enrolled in public daycare centers in Brazil and explored the relationships between gender and age to better understand and explain the results. Our main findings suggest gender and age differences. Girls outperformed boys in communication and personal-social skills, and the children's scores improved as they got older. The absence of an interaction between age and gender indicates that the two main effects occurred independently of each other and can be treated separately (Gelman et al., 2004).

With regard to communication skills, similar results have been reported by others (Catino et al., 2017; Filgueiras et al., 2013). From birth to the first year of life, female infants present stronger social orientation responses than male infants, with more interest in human faces, a greater amount of eye contact, and more accurate imitative abilities. Females exhibit activation of the left prefrontal cortex more often than males, suggesting a greater degree of executive processing and language-based decoding, with greater activation of mesolimbic regions. These abilities remain stable until later in life and are especially related to peer interactions and different interaction and communication styles (Wang et al., 2012).

Personal-social skills contribute to young children's self-help and adaption skills and include different socially acceptable behaviors that enable children to enter interpersonal relationships and interact with others and their environment. Other studies have shown that girls are, on average, more socially competent than boys. One explanation for this might lie in the fact that boys are more physically active, engage in more risk-taking behavior and rough-and-tumble play, and exhibit more anger and aggression toward peers than girls. Girls engage in more dyadic play and prefer the company of their mostly female preschool teachers (Vahedi, Farrokhi, & Farajian, 2012). Because the evaluation was performed by teachers, this process could have influenced the evaluation process. Numerous empirical studies have established that teachers perceive girls as more socially skilled than boys (O'Brennan, Bradshaw, & Furlong, 2014), which could have influenced the teachers' ratings.

The present study has several limitations. One limitation is related to public daycare centers. Another limitation is related to data gathering. Frequently, children who are enrolled in Brazilian public daycare centers are in an economically disadvantaged situation, and this compromises further generalization for children

with different demographic characteristics. We could not form a control group to measure the effect of the daycare facility itself on child development, which prevents us from claiming causal relationships between variables. We emphasize the both limitations are intrinsic to ethical standards and related to our research question.

One of the most important predictors of adult life satisfaction is emotional health in childhood. Monitoring this process enables the prevention and early detection of diseases and developmental disturbances in infancy and childhood (Campbell et al., 2014). Early identification and intervention lead to better social emotional and academic outcomes.

Conclusion

The present study investigated the development of children who were enrolled in public daycare centers in Brazil and assessed age and year differences and interactions based on ASQ-BR scores. The findings suggest that children who are enrolled in Brazilian public daycare centers, despite environmental adversity, are not at risk of serious developmental delays.

ARTICLE 5

Anunciação L, Chieh-Yu C, Squires J, Landeira-Fernandez J (2018) Screening for Social and Emotional Delays in Young Children Who Live in Poverty: A Brazilian Example. *J Child Dev Disord*. 4:5. doi: 10.4172/2472-1786.100055

Status: Published

Screening for Social and Emotional Delays in Young Children Who Live in Poverty: A Brazilian Example

Abstract: Emotional and social competence are notable predictors of future mental health outcomes. Studies have shown that poverty can negatively affect a child's development in several ways. In the Brazilian educational system, no direct payment is required to enroll children in public daycare centers. However, many daycare centers are in impoverished urban areas where the rates of violation of children's rights are still high. This situation is concerning because of the possible impact on children's development. The aim of this work was to investigate latent growth in 6,530 three- to four-year-old children who were enrolled in public daycare centers in the city of Rio de Janeiro in 2011 and 2012. We used a modified version of the Ages and Stages Questionnaire: Social and Emotional (ASQ:SE), in which 21 items across the questionnaires were retained. Latent Growth Modeling was performed by constraining intercepts of the repeated measures to one, and the slope's loadings corresponded to the study's time scale (in our case, 0 for age 3 and 1 for age 4). The intercept and slope results were significant ($p < .001$) and positive, indicating variability in the individuals' starting points. Consistent with these results, the scores increased as the children got older. Our findings suggest that children who are enrolled in Brazilian public daycare centers are achieving the expected emotional or social milestones that are appropriate for their age.

Keywords: Child development, Longitudinal research, Latent Growth Modeling, Ages and Stages Questionnaires, Daycare centers; Psychometrics

Introduction

Social and emotional domains are important predictors of mental health and the development of cognitive abilities and executive functions, including attention, working memory, and inhibitory control. Research suggests that deficits in emotion regulation and social competence are linked to greater levels of behavioral problems, difficulties with peers, and later psychopathology (Blandon, Calkins, & Keane, 2010; Kalvin et al., 2016). The development of both emotion regulation and social competence are also related to environmental characteristics (Gupta, de Wit, & McKeown, 2007).

In this direction, there is a vast international literature on the nature and extent of child poverty and a growing body of evidence on the consequences of child poverty: children who grow up in low-income environments face considerable barriers to healthy development and are more likely to be exposed to multiple environmental hazards, such as violence, crime, and drug abuse [(2)]. Low-income parents are often overwhelmed by reduced self-esteem, depression, and a sense of powerlessness and incapacity to cope—feelings that can get passed along to their children in the form of insufficient nurturing, negativity, and a overall failure to focus on children's needs (Gupta et al., 2007).

In Brazil, a governmental program began in 2010-2012 to assess the development of children who were enrolled in public daycare centers, and during 2011 and 2012, emotional and social aspects of child development were also assessed (Anunciacao, Squires, & Landeira-Fernandez, 2018). In public daycare centers, because no direct payment is required, mostly children come from low-income/very poor families. This program was highly influenced by evidence from the United States, where evaluations of early childhood education programs

demonstrated long-term impacts on a wide range of outcomes, including scholastic achievement, poverty, and criminal behavior.

The Brazilian program was unfortunately stopped in 2013, and few results are known about its outcomes (Anunciacao et al., 2018). That said, the present study focuses on describing and discussing child development based on the results that were obtained during this period.

Material and methods

The participants included a total of 6,530 children (52% boys and 48% girls) who were enrolled in 357 different daycare centers in Rio de Janeiro, Brazil. They were assessed by caregivers or teachers across 2 years (2011, when they were 3 years old; 2012, when they were 4 years old) using the Brazilian version of the Ages and Stages Questionnaires: Social and Emotional (ASQ:SE) [(4)]. More information about this procedure can be found in other publications [(5)].

We used 21 items across both 36- and 48-month ASQ:SE intervals to accommodate the ages of our preschool population. In the ASQ:SE traditional scoring system, higher scores indicate a risk for emotional and social problems. In this study, the system was reversed to reflect typical development. Because of that, we coded with "2" when participant checked the column of "Often or Always" for positive items or "Rarely or Never" for problematic items; "1", when the column "Sometimes" was checked; "0" when participant checked the column of "Rarely or Never" for positive items or "Often or Always" for problematic items (See Table 21).

Table 21. Twenty-one item ASQ:SE.

Domain	Content	Item (3 years)	Item (4 years)
Emotional	Child moves from one activity to the next	8	20
	Child settles down after exciting activity	7	7
	Child hurts self on purpose	22	23
	Child does what you ask	11	13
	Child cries, screams, or has tantrums for long periods	19	8
	Child tries to hurt other children	29	31
	Child calms down within 15 minutes	5	4
	Child destroys or damage things on purpose	24	25
	Child seems more active than other children of his/her age	12	16
	Child does things over and over and can't seem to stop	21	22
	Child sleeps at least 8 hours in a 24-hour period	16	15
	Child stays away from dangerous things	23	26
Social	Child can name a friend	26	27
	Child uses words to tell you what he/she wants	17	17
	Child uses words to describe his/her feelings	25	19
	Child plays/talk with adults he/she knows well	3	3
	Child is interested in things around her (people, toys)	10	9
	Child likes to play with other children	28	30
	Child looks at you when you talk	1	1
	Child seems happy	9	14
	Child likes to be hugged or cuddled	2	5

Ethical approval for this study was granted by the Ethical Committee of the Pontifical Catholic University of Rio de Janeiro. The data were analyzed using R 3.4 (R Development Core Team, 2016) and MPLUS V.8 (Muthén & Muthén, 2017) software.

Results and discussion

To be able to describe and compare the results that were obtained with the ASQ:SE across time points and to avoid potential interpretation bias, we checked

whether both versions that were used were statistically equivalent. We explored measurement invariance by verifying differences in practical fits, such as Confirmatory Fit Index (CFI) and Tucker Lewis Index (TLI) instead of checking a non-significant *p* value. According to the standard approach, the more constrained model is preferred only when the χ^2 test results in a nonsignificant *p* value ($p \geq .05$). The χ^2 value is inflated, however, when using large sample sizes (Widaman, Ferrer, & Conger, 2010).

Table 22. Fit indices of measurement invariance.

				χ^2					
	df	χ^2	$\Delta\chi^2$	df	*p*	CFI	RMSEA	ΔCFI	ΔRMSEA
Configural	376	6136				.923	.048	NA	NA
Loadings	395	6439.8	104.295	19	< .001	.926	.046	.003	.002
Intercepts	414	6446.5	30.949	19	.04	.924	.046	.003	0
Means	416	6882.2	312.256	2	< .01	.918	.047	.005	.001

As shown in Table 22, the results indicated longitudinal measurement equivalence. To assess changes in behavior as a function of the children's age, we performed Latent Growth Modeling. The intercept was constrained at one and the slope was constrained across time points (in our case, 0 for age 3 and 1 for age 4). Intercepts (initial status) and slopes (rate of change) are considered latent variables because they cannot be directly observed and represent aspects of change. Recent studies have shown that this approach is well suited to remove the effects of measurement error that might exist in predictors or outcomes (Curran, Obeidat, & Losardo, 2010).

The paths coefficients from both domains were: slope with intercept = .035, p = .019; slope mean = .384, p < .01; intercept variance = .267, p < .01, slope variance = .059, p < .01, (Figure 1 presents the standardized results). Once the difference between the two-time points was significant, we decided to report raw scores to facilitate understanding of the results and to compute the Cohen`s d effect size (Table 23).

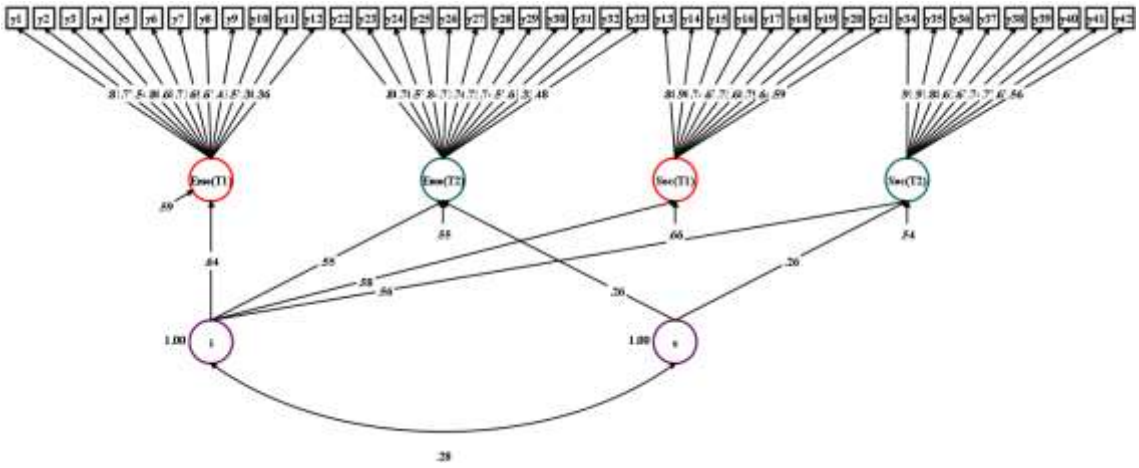

Figure 10 Standardized model results

Table 23. Descriptive results and effect size (Cohen's *d*).

		Female (*n = 3,095*)	Male (*n = 3,435*)	
		M (SD)	M (SD)	*Effect size*
Emotional	3 years	21.4 (2.92)	20.5 (3.51)	0.28
	4 years	22.3 (2.44)	21.4 (3.27)	0.31
	Effect size	0.33	0.27	
Social	3 years	16.8 (2.07)	16.4 (2.44)	0.18
	4 years	17.3 (1.59)	17 (1.95)	0.17
	Effect size	0.27	0.27	
Total	3 years	38.2 (3.9)	36.9 (4.68)	0.30
	4 years	39.7 (3.18)	38.4 (4.25)	0.35
	Effect size	0.42	0.34	

All scores increased as the children got older. Females tended to have higher scores on all domains of the ASQ:SE. If we assumed that higher scores

were associated with lower developmental risk for later social and emotional difficulty, then our findings are consistent with the literature (Blandon et al., 2010; Kalvin et al., 2016).

Other studies have shown that girls are, on average, more socially competent than boys. Externalizing behaviors (e.g., hyperactivity) are more common among boys than among girls. One explanation for this might lie in the fact that boys are more physically active, engage in more risk-taking behavior and rough-and-tumble play, and exhibit more anger and aggression toward peers than girls (Vahedi et al., 2012). When adults, the odds of having a psychological condition are higher in women (Riecher-Rössler, 2017).

Conclusion

The purpose of present study was to describe and discuss child development using data gathered with Brazilian children enrolled in public daycare centers. We used a longitudinal version of ASQ:SE with 21 items across both 36 and 48-month and assessed the changes in behavior using a Latent Growth Modeling.

Two important features in our study are 1. the items used on ASQ:SE and 2. the social conditions of the Brazilian public daycares, where data were gathered. First, since the items were the same, we could check the children's latent growth. Second, mostly of these children are growing up in low-income families, and because poverty is not exclusive to developing countries, our results can be useful for shedding light on child development in a challenging environment.

In our results, the intercept and slope variances were significant and positive, indicating variability in the individuals' starting points. In the same direction, the raw scores increased as the children got older. Our findings suggest

that children who are enrolled in Brazilian public daycare centers are achieving the expected emotional or social milestones that are appropriate for their age. Despite growing up in low income households, these children appear to be gaining social and emotional competence and performing well in their preschool environments.

Although there are some limitations (e.g., observational design with no control group), we believe that these results can be generalized to similar populations who live in poverty. In similar direction, as long the evaluation of the quality of early care and education is part of political agenda and expand the current understanding of public services, this study is also relevant (Ceglowski & Bacigalupa, 2002).

Longitudinal studies have long played a critically important role in developmental psychology and pediatric medicine, and these designs are becoming increasingly relevant to contemporary research. If researchers are able to estimate intra-individual patterns of changes over time, then they may better understand developmental trajectories of children and improve outcomes through early and targeted intervention.

Finally, we all agree that other studies are necessary to further document our findings, to address new questions about child development and to nurture an 'evidence-based policy-making' scenario using of statistics and statistical thinking throughout government decisions. Currently, new studies are being conducted to explore these issues.

5. General discussion

The two objectives of this thesis were (1) to investigate the main psychometric properties of the ASQ:SE using both Classical Test Theory methodology and Item Response Theory (IRT) and (2) to present an overview of child development in daycare services using the data that were gathered from this longitudinal project.

The thesis includes a theoretical background, in addition to five articles. The overall results provide evidence of the validity, reliability, and utility of the ASQ:SE. The fit statistics supported the hypothesized two-factor structure (emotional and social). The main findings generally allow concluding significant gender differences in ASQ and ASQ:SE outcomes.

Females had higher scores than males, and all scores increased as the children got older. Some of these results are well known. Other studies have shown that girls are, on average, more socially competent than boys. Externalizing behaviors, such as hyperactivity, are more common among boys than among girls. One explanation for this might lie in the fact that boys are more physically active, engage in more risk-taking behavior and rough-and-tumble play, and exhibit more anger and aggression toward peers than girls.

Developmental milestones include behaviors that demonstrate a child's growth in a number of areas, and they are established based on what most children can do at a certain age. All children develop at their own pace, but when children do not reach the milestones that are expected for their age, this could be an early warning sign of a possibly important developmental delay. Because of this, one important finding suggests that children who are enrolled in Brazilian public daycare centers, despite environmental adversity, are not at risk of serious developmental delays.

The present thesis has some limitations. One limitation is related to public daycare centers. Another limitation is related to data gathering. Frequently, children who are enrolled in Brazilian public daycare centers are in an economically disadvantaged situation, and this compromises further generalization to children with different demographic characteristics. It was also not possible to compare the results with a control group, which prevents us from claiming causal relationships between variables. Furthermore, as noted by Thurstone (1928), a measuring instrument must not be seriously affected in its measuring function by the object of

measurement. For many reasons, if these children were evaluated using a different psychometric tool, then the results could be different.

Overall, these findings may contribute to advances in three areas. First, preschool teachers and others who are interested in preventing developmental delays can apply the ASQ and ASQ:SE as an easy, contextually adapted, valid, and reliable psychometric tool. Second, public policies may benefit from the findings of this thesis. The ASQ system can be used to establish indicators to monitor early childhood development. Third, this thesis can contribute to future studies of the impact of non-cognitive skills on multiple outcomes, such as income and academic performance.

6. References

AERA, APA, & NCME. (2014). Standards for educational and psychological testing. In *American Educational Research Association*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:NCME+(1999)+Standards+for+educational+and+psychological+testing#0

Anunciacao, L., Squires, J., & Landeira-Fernandez, J. (2018). A Longitudinal Study of Child Development in Children Enrolled in Brazilian Public Daycare Centers. *Global Journal of Educational Studies*, *4*(1), 31. http://doi.org/10.5296/gjes.v4i1.12555

Apple, F. S. (2005). Validation of the 99th Percentile Cutoff Independent of Assay Imprecision (CV) for Cardiac Troponin Monitoring for Ruling Out Myocardial Infarction. *Clinical Chemistry*, *51*(11), 2198–2200. http://doi.org/10.1373/clinchem.2005.052886

Arias, V. B., Nuñez, D. E., Martínez-Molina, A., Ponce, F. P., Arias, B., Cooper, J., … Thissen, D. (2016). Hierarchy and Psychometric Properties of ADHD Symptoms in Spanish Children: An Application of the Graded Response Model. *PLOS ONE*, *11*(10), e0164474. http://doi.org/10.1371/journal.pone.0164474

Armijo, I., Schonhaut, L., & Cordero, M. (2015). Validation of the Chilean version of the Ages and Stages Questionnaire (ASQ-CL) in Community Health Settings. *Early Human Development*, *91*(12), 671–676. http://doi.org/10.1016/j.earlhumdev.2015.10.001

Baker, F. B. (2001). *The Basics of Item Response Theory*. *ERIC Clearinghouse on Assessment and Evaluation*. http://doi.org/10.1111/j.1365-2362.2010.02362.x

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824. http://doi.org/10.1016/j.paid.2006.09.018

Barros, R. P., Coutinho, D., & Mendonça, R. (2016). *Monitoramento e avaliação: desenhando e implementando programas de promoção do desenvolvimento infantil com base em evidências*. Brasília. Retrieved from https://www12.senado.leg.br/institucional/programas/primeira-infancia/pdf-arquivos/avancos-do-marco-legal-da-primeira-infancia

Blandon, A. Y., Calkins, S. D., & Keane, S. P. (2010). Predicting emotional and social competence during early childhood from toddler risk and maternal behavior. *Development and Psychopathology*, *22*(1), 119. http://doi.org/10.1017/S0954579409990307

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. http://doi.org/10.1037/0033-295X.110.2.203

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*(4), 1061–1071. http://doi.org/10.1037/0033-295X.111.4.1061

Brasil. Constituição da República Federativa do Brasil de 1988, Imprensa Nacional 1–139 (1988). Retrieved from http://www.planalto.gov.br/ccivil_03/constituicao/ConstituicaoCompilado.htm

Brasil. (2012). *Estatuto da criança e do adolescente*. *Edições Câmara*. http://doi.org/10.1590/0104-4060.188

BRASIL, M. da E. e C. (2011). *Monitoramento do uso dos indicadores da*

qualidade na educação infantil. Brasília.

Bricker, D., Allen, D., Clifford, J., Pretti-Frontczak, K., Slentz, K., & Squires, J. (2008). *The relationship between the Ages & Stages Questionnaires® (ASQ) and the Assessment, Evaluation, and Programming System for Infants and Children (AEPS®) (EMRG White Paper No. 1)*. Eugene, OR, OR: Early Intervention Management and Research Group (EMRG). Retrieved from http://aepslinkedsystem.com/pdf/White_Paper_ Relationship_Between_ASQ_and_AEPS_Final_1108.pdf

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children*. http://doi.org/10.4314/ai.v32i1.22297

Brown, C. M., Copeland, K. A., Sucharew, H., & Kahn, R. S. (2012). Social-emotional problems in preschool-aged children: opportunities for prevention and early intervention. *Archives of Pediatrics & Adolescent Medicine*, *166*(10), 926–32. http://doi.org/10.1001/archpediatrics.2012.793

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Reaearch*. *Journal of Chemical Information and Modeling*. http://doi.org/10.1017/CBO9781107415324.004

Browne, M. W. (2000). Psychometrics. *Journal of the American Statistical Association*, *95*(450), 661–665. http://doi.org/10.1080/01621459.2000.10474246

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-Term Memory, Working Memory, and Executive Functioning in Preschoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Developmental Neuropsychology*, *33*(3), 205–228. http://doi.org/10.1080/87565640801982312

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*. http://doi.org/10.18637/jss.v080.i01

Cai, L., & Hansen, M. (2013). Limited-information goodness-of-fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 245–276. http://doi.org/10.1111/j.2044-8317.2012.02050.x

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science*, *343*(6178), 1478–1485. http://doi.org/10.1126/science.1248429

Campos, M. M., Esposito, Y. L., Bhering, E., Gimenes, N., & Abuchaim, B. (2011). A qualidade da educação infantil: um estudo em seis capitais brasileiras. *Cadernos de Pesquisa*, *41*(142), 20–54. http://doi.org/http://dx.doi.org/10.1590/S0100-15742011000100003

Caprara, G. V., & Cervone, D. (2000). *Personality: Determinants, Dynamics, and Potentials*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511812767

Carey, W. B. (2001). From Neurons to Neighborhoods: The Science of Early Childhood Development. *Journal of Developmental & Behavioral Pediatrics*, *22*(2), 144–145. http://doi.org/10.1097/00004703-200104000-00011

Carlson, S. M. (2005). Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology*, *28*(2), 595–616. http://doi.org/10.1207/s15326942dn2802_3

Castro, M. A. de, Baltar, V. T., Selem, S. S. de C., Marchioni, D. M. L., &

Fisberg, R. M. (2015). Empirically derived dietary patterns: interpretability and construct validity according to different factor rotation methods. *Cadernos de Saúde Pública*, *31*(2), 298–310. http://doi.org/10.1590/0102-311X00070814

Catino, E., Di Trani, M., Giovannone, F., Manti, F., Nunziata, L., Piccari, F., … Sogos, C. (2017). Screening for Developmental Disorders in 3- and 4-Year-Old Italian Children: A Preliminary Study. *Frontiers in Pediatrics*, *5*. http://doi.org/10.3389/fped.2017.00181

Cattell, J. M. (1928). Early Psychological Laboratories. *Science*, *67*(1744), 543–548. http://doi.org/10.1126/science.67.1744.543

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. http://doi.org/10.1207/s15327906mbr0102_10

Ceglowski, D., & Bacigalupa, C. (2002). Four Perspectives on Child Care Quality. *Early Childhood Education Journal*, *30*(2), 87–92. http://doi.org/10.1023/A:1021245017431

Chen, C.-Y., Xie, H., Filgueiras, A., Squires, J., Anunciação, L., & Landeira-Fernandez, J. (2017). Examining the Psychometric Properties of the Brazilian Ages & Stages Questionnaires-Social-Emotional: Use in Public Child Daycare Centers in Brazil. *Journal of Child and Family Studies*. http://doi.org/10.1007/s10826-017-0770-0

Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research*, *18*(4), 447–460. http://doi.org/10.1007/s11136-009-9464-4

Coste, J., & Pouchot, J. (2003). A grey zone for quantitative diagnostic and screening tests. *International Journal of Epidemiology*, *32*(2), 304–313. http://doi.org/10.1093/ije/dyg054

Costin, C. (n.d.). Abertura: Políticas inovadoras de atenção à primeira infância - Claud…. Retrieved January 24, 2017, from http://pt.slideshare.net/saepr/seminrio-sae-2710

Cousineau, D. (2007). The rise of quantitative methods in psychology. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 1–3. http://doi.org/10.20982/tqmp.01.1.p001

Craig, K. (2017). The History of Psychometrics. In *Psychometric Testing* (pp. 1–14). Wiley. http://doi.org/10.1002/9781119183020.ch1

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development*, *11*(2), 121–136. http://doi.org/10.1080/15248371003699969

Darling-Churchill, K. E., & Lippman, L. (2016). Early childhood social and emotional development: Advancing the field of measurement. *Journal of Applied Developmental Psychology*, *45*, 1–7. http://doi.org/10.1016/j.appdev.2016.02.002

DeVellis, R. F. (1991). Guidelines in Scale Development. *Scale Development*, 51–60. http://doi.org/10.1017/CBO9781107415324.004

Ding, P. (2014). Bayesian robust inference of sample selection using selection-models. *Journal of Multivariate Analysis*, *124*, 451–464. http://doi.org/10.1016/j.jmva.2013.11.014

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222. http://doi.org/10.1186/2193-1801-2-222

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, *25*(2), 115–152. http://doi.org/10.1023/A:1024654026646

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self??? and Task Perceptions during Elementary School. *Child Development*, *64*(3), 830–847. http://doi.org/10.1111/j.1467-8624.1993.tb02946.x

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. http://doi.org/10.1037/1082-989X.5.2.155

Elafros, M. A., Bowles, R. P., Atadzhanov, M., Mbewe, E., Haworth, A., Chomba, E., & Birbeck, G. L. (2015). Reexamining epilepsy-associated stigma: validation of the Stigma Scale of Epilepsy in Zambia. *Quality of Life Research*, *24*(6), 1483–1489. http://doi.org/10.1007/s11136-014-0868-4

Elwick, A. (2018). New forms of government school provision – an international comparison. *Journal of Education Policy*, *33*(2), 206–225. http://doi.org/10.1080/02680939.2017.1329551

Ertem, I. O., Dogan, D. G., Gok, C. G., Kizilates, S. U., Caliskan, A., Atay, G., … Cicchetti, D. V. (2008). A Guide for Monitoring Child Development in Low- and Middle-Income Countries. *PEDIATRICS*, *121*(3), e581–e589. http://doi.org/10.1542/peds.2007-1771

Ferguson, B. (1990). Modern psychometrics. The science of psychological assessment. *Journal of Psychosomatic Research*, *34*(5), 598. http://doi.org/10.1016/0022-3999(90)90043-4

Fernald, L. C. H., Prado, E., Kariger, P., & Raikes, A. (2017). *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries*. Washington, DC.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psicológica*, *21*, 301–323.

Filgueiras, A., Pires, P., & Landeira-Fernandez, J. (2014). Screening Measures Used in Child Daycare Centers: A 15-Years Systematic Review. *Psychology*, *5*(19), 2109–2119. http://doi.org/10.4236/psych.2014.519213

Filgueiras, A., Pires, P., Maissonette, S., & Landeira-Fernandez, J. (2013). Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. *Early Human Development*, *89*(8), 561–576. http://doi.org/10.1016/j.earlhumdev.2013.02.005

Finch, W. H. (2011). A Comparison of Factor Rotation Methods for Dichotomous Data. *Journal of Modern Applied Statistical Methods*, *10*(2), 549–570. http://doi.org/10.22237/jmasm/1320120780

Fioravanti-bastos, A. C. M., Filgueiras, A., Lucia, M., & Moura, S. De. (2016). por profissionais de Educação Infantil Evaluation of the Ages and Stages Childhood professionals. *Estudos de Psicologia (Campinas)*, *33*(2), 293–302. http://doi.org/10.1590/1982-02752016000200011

Fonseca, F. F., Sena, R. K. R., Santos, R. L. A. dos, Dias, O. V., & Costa, S. de M. (2013). As vulnerabilidades na infância e adolescência e as políticas públicas brasileiras de intervenção. *Revista Paulista de Pediatria*, *31*(2), 258–264. http://doi.org/10.1590/S0103-05822013000200019

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics : an introduction*. Retrieved from http://catdir.loc.gov/catdir/enhancements/fy0808/2007016663-b.html

Gedler, G. R. (2004). Behavioral, social and emotional assessment of children and adolescents. *Psychology in the Schools*, *41*(5), 585–587. Retrieved from http://ezproxy.umsl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=12762513&site=ehost-live&scope=site

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis*. *Chapman Texts in Statistical Science Series*. http://doi.org/10.1007/s13398-014-0173-7.2

Goldstein, S., & Brooks, R. B. (2013). *Handbook of Resilience in Children*. (S. Goldstein & R. B. Brooks, Eds.). Boston, MA, MA: Springer US. http://doi.org/10.1007/978-1-4614-3661-4

GPH Panel of Experts. (2018). *IGNOU ACC-1 Organizing Child Care Services*. Gullybaba Publishing House Pvt. Ltd.

Gruendel, J. M. (2015). *When Brain Science Meets Public Policy: Institute for Child Success*.

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29–45. http://doi.org/10.1027/1901-2276.61.2.29

Gunzler, D. D., & Morris, N. (2015). A tutorial on structural equation modeling for analysis of overlapping symptoms in co-occurring conditions using MPlus. *Statistics in Medicine*, *34*(24), 3246–3280. http://doi.org/10.1002/sim.6541

Gupta, R. P.-S., de Wit, M. L., & McKeown, D. (2007). The impact of poverty on

the current and future health status of children. *Paediatrics & Child Health*, *12*(8), 667–672. http://doi.org/10.1093/pch/12.8.667

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology*, *45*, 8–18. http://doi.org/10.1016/j.appdev.2016.02.003

Heo, K. H., & Squires, J. (2012). Cultural adaptation of a parent completed social emotional screening instrument for young children: Ages and stages questionnaire-social emotional. *Early Human Development*, *88*(3), 151–158. http://doi.org/10.1016/j.earlhumdev.2011.07.019

Hilsenroth, M. J., Segal, D. L., & Hersen, M. (2003). *Comprehensive handbook of psychological assessment*. *Assessment* (Vol. 2). http://doi.org/10.1002/9780471726753

Jannuzzi, P. de M. (2002). Indicadores Sociais na Formulação e Avaliação de Políticas Públicas. *http://unesav.com.br/ckfinder/userfiles/files/INDICADORES-SOCIAIS-JANUZZI.pdf*. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf%5Cnhttp://smectamandare.hd1.com.br/sociais.pdf

Jee, S. H., Conn, A.-M., Szilagyi, P. G., Blumkin, A., Baldwin, C. D., & Szilagyi, M. A. (2010). Identification of Social-Emotional Problems among Young Children in Foster Care. *Journal of Child Psychology and Psychiatry*, *51*(12), 1351–1358. http://doi.org/10.1111/j.1469-7610.2010.02315.x

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. *American Journal of Public*

Health, *105*(11), 2283–2290. http://doi.org/10.2105/AJPH.2015.302630

Jones, S. M., & Kahn, J. (2017). *The evidence base for how we learn: Supporting students' social, emotional, and academic development. Consensus statements of evidence from the Council of Distinguished Scientists*.

Jones, L. V., & Thissen, D. (2007). A History and Overview of Psychometrics. *Handbook of Statistics*, *26*(6), 1–27. http://doi.org/10.1016/S0169-7161(06)26001-2

Kalvin, C. B., Bierman, K. L., & Gatzke-Kopp, L. M. (2016). Emotional Reactivity, Behavior Problems, and Social Adjustment at School Entry in a High-risk Sample. *Journal of Abnormal Child Psychology*, *44*(8), 1527–1541. http://doi.org/10.1007/s10802-016-0139-7

Kose, I. A., & Demirtasli, N. C. (2012). Comparison of Unidimensional and Multidimensional Models Based on Item Response Theory in Terms of Both Variables of Test Length and Sample Size. *Procedia - Social and Behavioral Sciences*, *46*, 135–140. http://doi.org/10.1016/j.sbspro.2012.05.082

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312. http://doi.org/10.1177/1745691611406925

Kucuker, S., Kapci, E. G., & Uslu, R. I. (2011). Evaluation of the Turkish Version of the "Ages and Stages Questionnaires. *Infants & Young Children*, *24*(2), 207–220. http://doi.org/10.1097/IYC.0b013e31820eae26

Lagerspetz, K. M. J., Björkqvist, K., & Peltonen, T. (1988). Is indirect aggression typical of females? gender differences in aggressiveness in 11??? to 12???year???old children. *Aggressive Behavior*, *14*(6), 403–414. http://doi.org/10.1002/1098-2337(1988)14:6<403::AID-

AB2480140602>3.0.CO;2-D

Le, D.-T. (2014). Applying item response theory modeling in educational research. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *75*(1).

Lee, N. R., Wallace, G. L., Raznahan, A., Clasen, L. S., & Giedd, J. N. (2014). Trail making test performance in youth varies as a function of anatomical coupling between the prefrontal cortex and distributed cortical regions. *Frontiers in Psychology*, *5*. http://doi.org/10.3389/fpsyg.2014.00496

Levinson, C. M. D. (2011). Social-Emotional Problems Among Low Income Preschool-Aged Children and Potential Factors Affecting Early Intervention. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=ucin1307126740

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., … Maris, G. (2018). An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models. *Multivariate Behavioral Research*, *53*(1), 15–35. http://doi.org/10.1080/00273171.2017.1379379

Maydeu-Olivares, A. (2013). Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 71–101. http://doi.org/10.1080/15366367.2013.831680

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin*, *115*(2), 300–307. http://doi.org/10.1037/0033-2909.115.2.300

Muthén, L., & Muthén, B. (2017). *Mplus user's guide (8th ed.)*. *Los Angeles: Author*. http://doi.org/10.13155/29825

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of*

Mathematical Psychology, *47*(1), 90–100. http://doi.org/10.1016/S0022-2496(02)00028-7

Novick, M. R. (1980). Statistics as psychometrics. *Psychometrika*, *45*(4), 411–424. http://doi.org/10.1007/BF02293605

Nowakowska, M. (1983). Chapter 2 Factor Analysis: Arbitrary Decisions within Mathematical Model (pp. 115–182). http://doi.org/10.1016/S0166-4115(08)62370-5

O'Brennan, L. M., Bradshaw, C. P., & Furlong, M. J. (2014). Influence of Classroom and School Climate on Teacher Perceptions of Student Problem Behavior. *School Mental Health*, *6*(2), 125–136. http://doi.org/10.1007/s12310-014-9118-8

Olsbjerg, M., & Christensen, K. B. (2015). Modeling local dependence in longitudinal IRT models. *Behavior Research Methods*, *47*(4), 1413–1424. http://doi.org/10.3758/s13428-014-0553-0

Papalia, D. E., Olds, S. W., & Feldman, R. D. (2008). *A Child's World: Infancy through Adolescence*. New York: McGraw-Hill.

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, *16*(2), 56–69. http://doi.org/10.17605/OSF.IO/TNRXV

Pollock, J., & Horrocks, S. (2008). *Evaluation of the Primary Mental Health Specialist (Infant Mental Health) Programme in North Bristol NHS Trust (NBT) and United Bristol Healthcare NHS Trust (UBHT)*.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the Optimal Number of Factors in Exploratory Factor Analysis: A Model Selection Perspective. *http://dx.doi.org/10.1080/00273171.2012.710386*.

R Development Core Team. (2016). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing Vienna Austria*, *0*, {ISBN} 3-900051-07-0. http://doi.org/10.1038/sj.hdy.6800737

Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling: Applications to typical performance assessment.*

Riecher-Rössler, A. (2017). Sex and gender differences in mental disorders. *The Lancet Psychiatry*, *4*(1), 8–9. http://doi.org/10.1016/S2215-0366(16)30348-0

Rio de Janeiro City. (2017). *Quick Guide to Public Policies*. Rio de Janeiro. Retrieved from http://www.rio.rj.gov.br/dlstatic/10112/4379008/4130265/EducacaoING1211.pdf

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. http://doi.org/10.1016/j.jmp.2012.08.001

Sabanathan, S., Wills, B., & Gladstone, M. (2015). Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Archives of Disease in Childhood*, *100*(5), 482–488. http://doi.org/10.1136/archdischild-2014-308114

Schimmenti, A., & Bifulco, A. (2015). Linking lack of care in childhood to anxiety disorders in emerging adulthood: the role of attachment styles. *Child and Adolescent Mental Health*, *20*(1), 41–48. http://doi.org/10.1111/camh.12051

Schonhaut, L., Armijo, I., Schonstedt, M., Alvarez, J., & Cordero, M. (2013). Validity of the Ages and Stages Questionnaires in Term and Preterm Infants. *PEDIATRICS*, *131*(5), e1468–e1474. http://doi.org/10.1542/peds.2012-3313

Sijtsma, K. (2013). Theory Development as a Precursor for Test Validity (pp. 267–274). Springer New York. http://doi.org/10.1007/978-1-4614-9348-8_17

Sijtsma, K., & Junker, B. W. (2006). Item response theory: Past performance, present developments, and future expectations. *Behaviormetrika*, *33*(1), 75–102. http://doi.org/10.2333/bhmk.33.75

Singh, A., Squires, J., Yeh, C., Heo, K., & Bian, H. (2016). Validity and reliability of the developmental assessment screening scale. *Journal of Family Medicine and Primary Care*, *5*(1), 124. http://doi.org/10.4103/2249-4863.184636

Singh, G. (2006). Determination of Cutoff Score for a Diagnostic Test. *The Internet Journal of Laboratory Medicine*, *2*(1). http://doi.org/10.5580/1aca

Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires[R], Third Edition (ASQ-3[TM]): A Parent-Completed Child-Monitoring System*. *Brookes Publishing Company*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED507532&site=ehost-live&scope=site%5Cnhttp://www.brookespublishing.com/store/books/squires-asq/index.htm

Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure *. *Early Childhood Research Quarterly*, *16*(4), 405–419. http://doi.org/10.1016/S0885-2006(01)00115-6

Squires, J., Bricker, D., & Potter, L. W. (1997). Revision of a parent-completed developmental screening tool: Ages and stages questionnaires. *Journal of*

Pediatric Psychology, *22*(3), 313–328. http://doi.org/10.1093/jpepsy/22.3.313

Stenberg, K., Axelson, H., Sheehan, P., Anderson, I., Gülmezoglu, A. M., Temmerman, M., … Bustreo, F. (2014). Advancing social and economic development by investing in women's and children's health: a new Global Investment Framework. *The Lancet*, *383*(9925), 1333–1354. http://doi.org/10.1016/S0140-6736(13)62231-X

Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*, *39*(3), 135–140.

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size - or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–82. http://doi.org/10.4300/JGME-D-12-00156.1

Tatagiba, A. P. (2010). Aspectos da política educacional carioca: trajetórias da educação infantil. *Ensaio: Avaliação E Políticas Públicas Em Educação*, *18*(67), 275–302. http://doi.org/10.1590/S0104-40362010000200006

Vahedi, S., Farrokhi, F., & Farajian, F. (2012). Social competence and behavior problems in preschool children. *Iranian Journal of Psychiatry*, *7*(3), 126–134.

van Heerden, A., Sen, D., Desmond, C., Louw, J., & Richter, L. (2017). App-Supported Promotion of Child Growth and Development by Community Health Workers in Kenya: Feasibility and Acceptability Study. *JMIR mHealth and uHealth*, *5*(12), e182. http://doi.org/10.2196/mhealth.6911

Wallander, J. L., Bann, C. M., Biasini, F. J., Goudar, S. S., Pasha, O., Chomba, E., … Carlo, W. A. (2014). Development of children at risk for adverse outcomes participating in early intervention in developing countries: a

randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *55*(11), 1251–1259. http://doi.org/10.1111/jcpp.12247

Wang, Y., Adamson, C., Yuan, W., Altaye, M., Rajagopal, A., Byars, A. W., & Holland, S. K. (2012). Sex differences in white matter development during adolescence: A DTI study. *Brain Research*, *1478*, 1–15. http://doi.org/10.1016/j.brainres.2012.08.038

Wei, M., Bian, X., Squires, J., Yao, G., Wang, X., Xie, H., … Chen, J. (2015). [Studies of the norm and psychometrical properties of the ages and stages questionnaires, third edition, with a Chinese national sample]. *Zhonghua Er Ke Za Zhi = Chinese Journal of Pediatrics*, *53*(12), 913–918. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26887546

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, *4*(1), 10–18. http://doi.org/10.1111/j.1750-8606.2009.00110.x

Zumbo, B. D., & Hubley, A. M. (Eds.). (2017). *Understanding and Investigating Response Processes in Validation Research* (Vol. 69). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-56129-5