How many infants are temperamentally difficult? Comparing norms from the Revised Infant Temperament Questionnaire to a population sample of UK infants

Shiau Yun Chong a,*, Catherine R Chittleborough a, Tess Gregory a, b, John W Lynch a, c, Lisa G Smithers a

a School of Population Health, University of Adelaide, Adelaide, 5005, Australia
b Telethon Kids Institute, The University of Western Australia, Perth, Australia
c School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

ARTICLE INFO

Article history:
Received 20 November 2014
Received in revised form 13 April 2015
Accepted 18 April 2015
Available online 25 May 2015

Abstract

The original norms for the Revised Infant Temperament Questionnaire (RITQ) were published in 1978 and were based on a small sample from the US. The aim of this study is to compare temperament scores from the original RITQ against scores from a large population-based cohort of infants from the UK. This study consists of 10,937 infants from the Avon Longitudinal Study of Parents and Children (ALSPAC) born between April 1991 and December 1992 in the southwest of England. Infant temperament at 6 months of age was reported by parents using the adapted RITQ. Responses were scored according to the RITQ manual and then categorized into temperament groups (easy, intermediate low, intermediate high, and difficult) using either the RITQ norms or norms derived from the data. The scores for each temperament subscale and the proportion of children in each temperament group were compared across the two methods. Subscale scores for the ALSPAC sample were higher (more “difficult”) than the RITQ norms for rhythmicity, approach, adaptability, intensity, and distractibility. When RITQ norms were applied, 24% infants were categorized as difficult and 25% as easy, compared with 15% difficult and 38% easy when ALSPAC norms were used. There are discrepancies between RITQ norms and the ALSPAC norms which resulted in differences in the distribution of temperament groups. There is a need to re-examine RITQ norms and categorization for use in primary care practice and contemporary population-based studies.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Infant temperament assessment is recommended for health practitioners including pediatricians, physicians, and pediatric nurse practitioners as part of their routine screenings. The Infant Temperament Questionnaire (ITQ) is one of the well-established tools for assessment of infant’s temperament. The ITQ was published in 1970 by the pediatrician,
Dr William Carey (Carey, 1970) based on results published by Thomas, Chess and colleagues from the New York Longitudinal Study (NYLS) (Thomas, Chess, Birch, Hertzig, & Korn, 1963). Thomas and Chess identified nine dimensions of temperament from extensive observations and qualitative interviews with the parents of 22 children in the NYLS and these form the nine subscales of temperament in the ITQ (activity, rhythmicity, adaptability, approach, intensity, mood, persistence, distractibility, and threshold). The ITQ was revised in 1978 by Carey and McDevitt, and the Revised Infant Temperament Questionnaire (RITQ) has shown moderate internal consistency (0.49 to 0.71 for subscales, 0.83 for composite) and good test-retest reliability (0.66 to 0.81) (Carey & McDevitt, 1978).

The RITQ is used extensively by health practitioners in primary care settings (Garstein, Bridgett, & Low, 2012; Hertzig & Snow, 1988; Younger, 2011) as it is clinically derived and is useful for identifying childhood clinical conditions such as behavioral problems (Carey & McDevitt, 2012; Stein, Plonsky, Zuckerman, & Carey, 2005; Younger, 2011). The concept of “goodness of fit” introduced by Thomas and Chess is particularly useful in clinical interventions to help parents understand the importance of the consonance between the child’s temperament and the expectations of the parents to the development of the child (Garstein et al., 2012). Parenting advice on how to manage the child’s behavior can then be given according to the temperament profile of the child (Andersen, 2000, 2002). For example, in Bright Futures in Practice: Mental Health, a set of pediatric guidelines for promoting socio-emotional wellbeing of children from birth through adolescence, the RITQ temperament subscales were explained and strategies on how to improve the “fit” for the children were provided (Jellinek, Patel, & Froehle, 2002). A number of parent education materials have also been developed based on the RITQ temperament subscales (Kurcinka, 1998; Neville & Williams, 2007).

The RITQ is also used in many large-scale longitudinal studies such as the Millennium Cohort Study (Pickett, Wood, Adamson, DeSouza, & Wakschlag, 2008), the National Institute of Child Health and Human Development Study of Early Child Care (Bradley & Corwyn, 2008), and the Helsinki Longitudinal Temperament Study (Martin, Wisenbaker, Baker, & Huttunen, 1997). Longitudinal studies have provided evidence that temperament is associated with later development of mental disorder (Lewis & Olsson, 2011; Sayal, Heron, Maughan, Rowe, & Ramchandani, 2013), behavioral problems (Prior, Sanson, Smart, & Oberklaid, 2000), as well as cognitive (Maziade, Côté, Boutin, & Bernier, 1987), language (Taylor, Christensen, Lawrence, Mitrou, & Zubrick, 2013), and academic performance outcomes (Saudino & Plomin, 2007). Assessing infant temperament in large-scale community or population samples helps identify infants who may be at increased risk for later cognitive, academic or behavioral problems and may assist policy makers to better target groups of children for interventions. For example, parenting programs may be considered to provide targeted support to families with temperamentally difficult children.

As part of the RITQ, a profile sheet provides means and standard deviations (SDs) for each subscale that can be used to identify an infant’s temperament profile (Carey & McDevitt, 1977). These means and SDs were derived from a standardization on 203 infants (104 boys and 99 girls) aged 4- to 8-months old, predominantly from middle-to-upper class US families in 1978 (Carey & McDevitt, 1978). When examining an infant’s temperament, clinicians compare the infant’s scores on each temperament subscale with the normative scores on the RITQ defined by the 1978 sample. Infants can also be categorized into different temperament groups (easy, intermediate low, intermediate high, and difficult) based on where their scores sit in relation to the normative sample. This categorization is then used for subsequent investigations and interventions, and parenting advice can be provided to parents according to the temperament profile of their child.

Some studies have used their own sample norms to categorize infants into temperament groups. For instance, in a study of 985 infants in the United States, infant temperament was categorized into three groups (easy, average, and difficult) using the study sample means and SDs as cut-offs (Bradley & Corwyn, 2008). Other studies have used the RITQ norms to categorize infants into temperament groups but have shown significant differences between the means and SDs observed in their sample and those established based on the 1978 sample. For example, in a study of 349 infants aged 4 to 8 months in Taiwan, RITQ items were translated into Chinese and then translated back into English. The study found that infants scored significantly higher than the RITQ standardization sample in approach, adaptability, mood, intensity, distractibility, and threshold (Hsu, Soong, Stigler, Hong, & Liang, 1981). Another study using a Japanese version of the RITQ had means that were higher than the RITQ means on all subscales except activity and threshold (Sasaki, Mizuno, Kaneko, Murase, & Honjo, 2006). However, the Japanese version of the RITQ has not been back translated, so it is difficult to tease apart whether the differences are due to translation or context-specific perceptions of difficult temperament. While using different means and SDs could potentially result in inaccurate identification of infant temperament, previous research using the RITQ has not addressed this issue.

As stated in the manual (Carey & McDevitt, 1977), RITQ norms published based on the standardization sample may not apply to other populations. Norms may vary across cultures and populations. However, to date, there are limited studies with large, representative samples that have published population-specific norms using the RITQ items. This is a problem for RITQ users because this could lead to misclassification of temperament if population-specific norms differ from the RITQ norms.

This study aims to compare the original norms (means and SDs) published for the RITQ using the 1978 sample (Carey & McDevitt, 1978) with norms derived from a large population sample of UK infants, and the resulting categorization of temperament (easy, intermediate low, intermediate high, and difficult) from these two different norms.