



COMPARATIVE EARLY INFANT MILESTONE DEVELOPMENT: MEASURING BABIES AND ENGAGING PARENTS, EAST AND WEST

Anthony R. Dickinson^{*} Elaine Zheng^{}**
Academic Research Laboratory, USA

Abstract

We report the first results of a research program which has followed more than 400 infant/parent cohorts who were enrolled in, and attended, continuous weekly/monthly developmental training programs conducted in Asia. Independent measures included regular use of a norm-standardised developmental assessment tool (ASQ, 2006), designed to reveal the developmental milestone attainment times for babies aged 6-33 months of age, inclusively. Results showed elevated scorings for all behaviours measured, across each of five core developmental domains, in comparison to the published norms for age-matched babies born in the USA ($p < 0.01$). More surprisingly, far higher numbers of 'Advanced' performing infants ($p < 0.001$), and far fewer scoring consistent with 'Delayed' performance levels ($p < 0.001$), were recorded after 10 months of developmental achievement training program interaction in four Asian cities. Encouraging caution in supporting our current reliance upon the published normative samples of North American baby developmental achievements for individual child comparisons (as provided by the test publishers), our results suggest there to be a quite different profile of normative developmental abilities observable in China. Our results indicate both higher and faster growing ability rate measures to be now detectable

Correspondence concerning this paper should be addressed to:

^{*} *Academic Research Laboratory: Gluck Rm 712E , 7/F , Witty Building, 1A - 1L Tong Choi Street, New Territories, HK (SAR) PRC and
Institute of Neuroscience (Chinese Academy of Sciences), 320 Yue Yang Rd., Shanghai, China, 200031, PRC. E-mail: td2202@hotmail.com*

^{**} *Academic Research Laboratory: Gluck Rm 712E , 7/F , Witty Building, 1A - 1L Tong Choi Street, New Territories, HK (SAR) PRC and
KB International Education (USA): Ste 104, Rodney Bldg, 3411 Silverside Rd., New Castle, Wilmington, DE, USA 19810*

across the standard range of infant developmental performance domain indicators, and thus the need for Asia-Pacific region-specific norm-tables.

Keywords: infant psychometrics, infant education, baby development, parenting, comparative psychology

Introduction

The literature concerned with the determination of comparative ability, and age-related milestone achievement performance measures in human babies and young infants reared in different parts of the world, has remained particularly elusive, and is certainly lacking in its use of any clear or consistent methodology. Indeed, with the notable exception of some ongoing research conducted with Asian-Indian baby populations within the context of ‘Healthy Baby Contests’ (Prajapati, 1998; Kulkarni & Mohan, 1999), there has to date been no coherent comparative analysis of baby developmental milestone achievements (of cross-cultural significance), which makes any systematic use of norm-based internationally-recognised, standardised testing tools. In order to address the need to involve the parents of the babies participating in this study (see the second of the two study objectives, below), our choice of a norm-referenced ‘typical ability’ test was necessarily limited (otherwise requiring significantly prohibitive testing times, and the presence of a certified professional clinical psychometrician). The outstanding exception presented itself in the form of an assessment test battery of measures known as the *Ages and Stages Questionnaire (ASQ-II)*, Squires & Bricker, 2006), a relatively simple, yet effective indicator of both high and low levels of a wide range of critical developmental abilities. This non-clinical test is frequently used as a parent self-report screening tool prior to the use of professional composite testing (e.g., Bayley-III), as we would use more formally in clinical, educational and psychological research settings (Dickinson, Hui, & Zheng, 2011). The *ASQ-II* norm-table comparisons as provided by the test Publisher are for North American (Caucasian and Hispanic) samples, and include separate entries for each of 5 ability achievement domains (Communication,

Gross Motor, Fine Motor, Problem Solving, and Socio-personal developmental milestones), with detailed norm-tables provided for each of ten 2-3 month age groupings (covering baby-infant ages from 4 to 40 Mths).

Once the individual baby behaviour measurement tool for use in this study had been identified, the authors then set about designing an integrated assessment program of research which would achieve three principle functions. Firstly, to provide a comparative longitudinal measure of individual baby/infant's rate(s) of developmental milestone achievements, using a standardised (norm-referenced) testing tool, which would allow comparison between the apparent ability attainments of newly born and developing Asian and American infants. Our second task was to provide an assessment tool which would both engage, and help educate, the parents of the babies who were being assessed. The third task was to be able (in real time) to provide critical developmental information concerning each individual baby's relative strengths and weaknesses, across multiple behavioural performance domains, for the purpose of creating/optimising student-centred training class activity design and curriculum planning.

Study objectives

Having satisfied each of the three objectives as set out above, it was the aim of the current paper to then specifically address the following two research questions: Firstly, (1) might the ASQ baby milestone test battery allow not only the parent to measure, but also become increasingly engaged in the process of monitoring, understanding and promoting their baby's growth? This hypothesis would be tested using analysis of parent's continuing participation rates, and anecdotal reports. Secondly, (2) whether we could usefully compare the growth trajectory and perceived abilities of Asian and Western babies? The *null hypothesis* predicts that there would be no significant differences ($p < 0.05$) found between any age or stage timings of reported developmental milestone achievements for the Asian baby cohort, when compared to the norm-referenced sample of North American babies, as published in the ASQ test manuals supplied.

Methods

Participants

The present study recruited (n=404) self-selecting, volunteer infant/parent cohorts who were each enrolled in, and attended, one of four continuous weekly developmental ‘parent and baby’ training program classes, held over the course of one year, and simultaneously running in four major Chinese cities (Dickinson & Zheng, 2012).

Instruments

Concurrent with their attendance at training program sessions, regular bi-monthly use of the norm-based ASQ-II developmental assessment tool (Squires & Bricker, 2006) was administered according to the publisher’s *Test User’s Manual* provided, for the purpose of revealing developmental milestone attainment times for all research-cohort babies from 8-33 months of age. The test papers were made available in the English and Chinese languages (both Simplified & Traditional forms in the case of the latter), with verbal translators competent with all three scripts present, and available at all times during the test administrations. Each participant parent completed one ASQ-II test battery every time their child attained a new age-appropriate level (as made available for months 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, and 36, respectively), the information from which was normally data-base collated the same day as that upon which the test administration was conducted. Parents were provided confidential feedback within three days concerning their own babies’ comparative performance across the five domains of Communication, Fine Motor, Gross Motor, Cognition & Learning, and Social-Personal dimensions of development (with reference to the ASQ-II test publisher’s norm-tables as provided, and using the aggregate means and standard deviations as presented in Table 3 on p.41 of the same test manual (Squires & Bricker, 2006), in assessing norm versus individual baby (experimental sample) comparisons. All test administrators were required to attend (and pass) a multi-day course of instruction (with examination), which was organised and supervised by the study’s principal investigator. All of the research study team (including the current authors) remained naïve as to the identities of any given baby infant’s scores, and gained access to the cohort-identity-coded data-base only once the 10 month test administration period was completed. Statistical result analyses

were conducted using SPSS software mounted on a portable laptop computer. All participants were later invited to attend one of several informal debriefing sessions held at their respective training locations after completion of the 12-month study period, at which the principal investigator presented an overview of the results of the study's findings, and invited questions or comments for discussion.

Results

The critical data for analysis were determined in accordance with the acquisition of more than 5000 ASQ test item responses as provided by the primary caregivers of individual babies (mostly, but not exclusively their mothers), and were collated over a continuous 10-12 month period of training program attendance, providing longitudinal indicators of each child's typical performance attainment across five distinct ability domains. Participant drop-out rates were below 5%. This result alone, together with other longitudinal anecdotal data to be reported elsewhere (Dickinson & Zheng, 2013, in prep), provided a clear positive and affirmative answer to our first research question and hypothesis (a 95% completion over one year being all the more remarkable when considering that Q-based responses even for a single response test-battery indicator are usually considered 'reliable' when returned at 25% target sample sizes (Curtin, Presser, & Singer, 2000)). The ASQ test analysis of quantitative comparative results showed elevated scorings to be recorded for *all* 'typical ability' behaviours measured, across each of the five domains (Communication, Gross Motor, Fine Motor, Problem Solving, and Socio-Personal skills), in contrast to the published norms for age-matched babies from the USA (aggregate mean scores [corrected using receiver-operator characteristic analysis (Squires & Bricker, 2006, App F)] = 55.8 vs 51.4, $p < 0.01$, representing close to a 10% difference between the Asian study cohort and North American baby norm populations, respectively). Although the mean scores for the experimental cohort in each of the developmental areas (and for each age range) were higher than the norm-indicated values (and despite the existence of a pre-determined ceiling limit of attainment, see Table 1 for typical example comparisons), more surprisingly, far higher numbers of 'Advanced'

performing infants (50%+ vs 18%, $p < 0.001$), and far fewer scoring consistent with ‘*Delayed*’ performance levels (1% vs 18%, $p < 0.001$) [for each of the Asian study cohort and North American baby norm populations, respectively], were found to be recorded after only 8 mths of their developmental achievements during training program interaction.

Table 1. Means and Standard deviations (SD) for each of 5 developmental areas: showing 3 different age intervals, typical of USA Norm Vs Asian study baby cohorts

Age Interval (Mths)	Developmental Area	Mean Score (SD) USA Norm	Mean Score (SD) Asian Cohort
8	Communication	53.5 (8.4)	56.2 (6.4)
	Gross Motor	50.5 (13.1)	57.4 (5.3)
	Fine Motor	54.4 (8.8)	56.3 (6.4)
	Problem solving	51.7 (9.7)	52.9 (7.2)
	Socio-personal	51.3 (10.4)	54.8 (6.6)
16	Communication	49.1 (12.3)	56.4 (7.1)
	Gross Motor	55.3 (11.5)	58.8 (4.3)
	Fine Motor	51.8 (10.6)	53.5 (6.9)
	Problem solving	49.7 (11.4)	54.6 (8.4)
	Socio-personal	48.5 (10.9)	57.3 (4.8)
24	Communication	49.5 (11.5)	58.3 (4.9)
	Gross Motor	54.4 (9.2)	55.7 (6.2)
	Fine Motor	52.8 (8.2)	57.1 (5.4)
	Problem solving	51.5 (9.3)	56.4 (6.7)
	Socio-personal	52.4 (8.4)	56.0 (5.6)

Looking at the data in more detail, figure 1 shows the percentage of babies reportedly showing *delayed* levels of behavioural achievement for their age, as determined by the test publisher’s norm tables. All ANOVA analyses were highly significant ($p < .001$) for all test comparisons conducted (Asian study cohort versus North American), with no significant differences found for any effects of city location, language of test administration, baby gender or respondent parent gender *within* the Asian study cohort sample, though there was a slight (but non-significant) effect of training centre attended ($p > 0.05 = ns$).

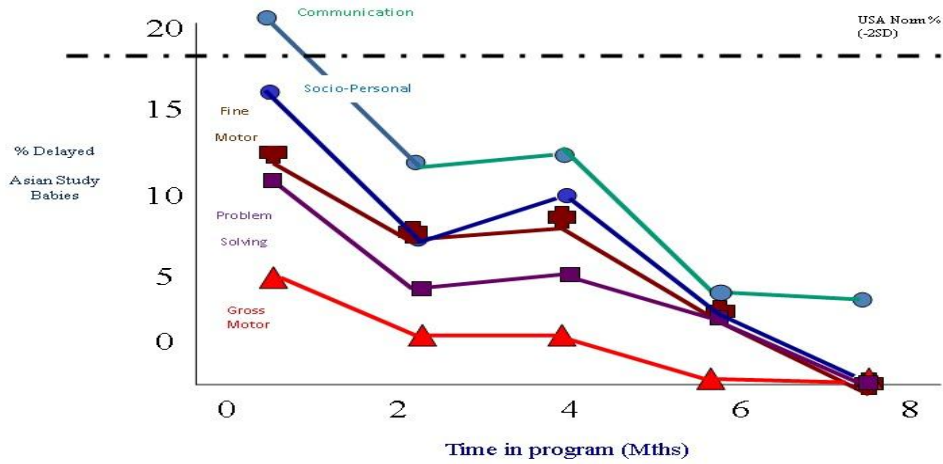


Figure 1. Percentage babies reportedly showing *Delayed* levels of ability

In clear contrast to the North American cohort normalised validity indicators for *developmental delay* categorisation (dot-dash line, -2SD, or minus two Standard Deviations from the mean), very rapid falls in the numbers of Asian study cohort babies identified as showing developmental delay were seen as they continued to attend their training sessions. Indeed, for four out of the five behavioural domains measured, none of the Asian study cohort babies were so categorized after 10 months attendance, in comparison to the 17% or so expected according to the North American cohort norms.

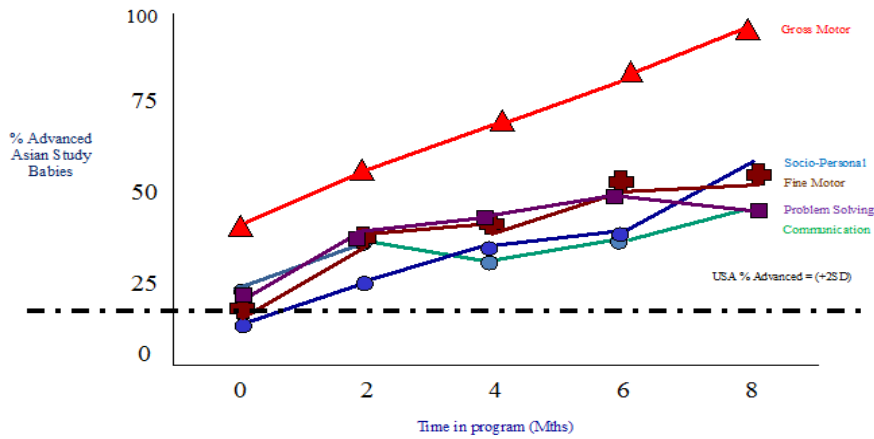


Figure 2. Percentage babies reportedly showing *Advanced* levels of ability

Similarly for indicators of *advanced development* with their ‘typical ability’ categorisations (see figure 2, the dot-dash line this time +2SD from the mean), this time very rapid *increases* in the numbers of Asian study cohort babies were identified as showing advanced development profiles as they continued to attend their training sessions. Indeed, for three out of the five behavioural domains measured, more than 50% of the Asian study cohort babies were reportedly showing advanced levels of behaviour development, in comparison to the 17% or so found to be performing above this level in the North American baby cohort population.

In case these surprisingly large differentials between Asian study cohort and North American ASQ attainments might be explained by our having pooled across ages of babies at assessment (they are collapsed across ages as shown in figures 1-2), figure 3 shows the same data plotted according to the actual ages of the babies at the time of their evaluations (rather than by their time in program attendance alone).

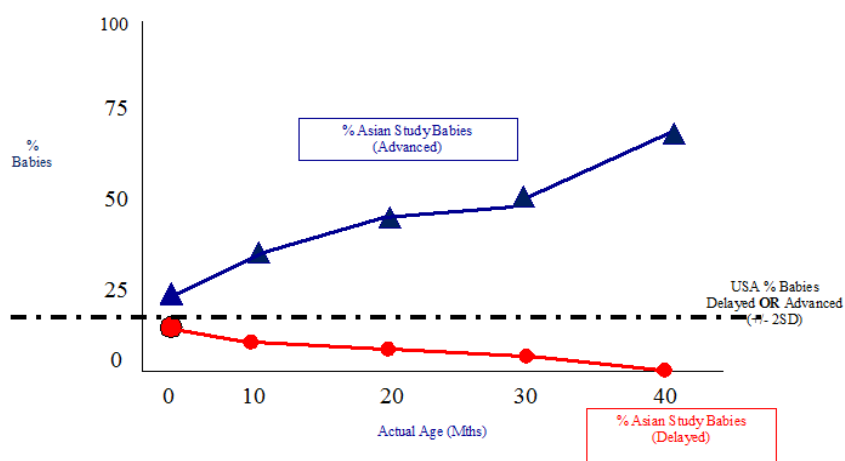


Figure 3. Percentage babies showing Advanced or Delayed levels of ability, by real age

Now controlling for absolute ages, however (independent of the time the parent-baby cohorts had spent in training), even stronger profiles are revealed when contrasting the reports of the Asian study cohort babies with

the American sample norms (with all ANOVA analyses showing differences to exist at even higher levels of significance ($p < .0005$)).

Discussion

Although the ASQ test is relatively simple to administer, it remains nonetheless an effective indicator of both high and low levels of a wide range of critical developmental abilities. This non-clinical test, though frequently used as a pre-screening tool prior to the use of professional composite testing referrals, it may also be usefully employed to enhance parental involvement in the monitoring, early education and scaffolding of their baby's developmental ability levels. Its use likewise as a tool to aid the design of specific training activities to optimise the growth of each individual child (once their individual strengths and weaknesses be identified), cannot now be underestimated.

With respect to the purely professional psychometric implications of our findings (see also comments by Bhat, Pardeshi, Kakrani, and Pratinidh, 2001), and in contrast to our current reliance upon the published normative samples of North American baby developmental achievements for individual child comparisons (as provided by the test publishers), our results suggest there to be a quite different profile of normative developmental ability to be found in Asia today. In particular, the current study suggests the existence of both higher, and faster growing ability rate measures, across a wide range of infant developmental performance domain indicators, at least in the case of our study's (mostly) Chinese-born babies. Whether or not these be taken to indicate the effects of the training program attended, confirmation of such a claim would require a matched cohort of comparative North American babies who were not only age-matched, but also undergoing such training activity interaction as was being made available to the Asian parent-baby cohort whose developmental behaviours were being measured in the current study.

As with the similar controversy concerned with comparative and cross-cultural IQ score data, some researchers may wish to argue that the current study results may support the search for the oft sought longitudinal study of

gifted and talented Asian children from developing infancy (who have otherwise rarely involved the identification of their unusual abilities before reaching 12 years of age (Dickens & Flynn, 2001), the few studies that *have* been published, having sought to do so using only standardised intelligence [IQ] tests of different kinds (Yung & Dickinson, 2008).

However, it is the current authors' conclusion that, training effects aside, this study's findings lead us to propose there to be an urgent need for the generation and dissemination of new Chinese/Asia-Pacific population-based developmental test norms, for the purposes of facilitating significant ASQ test interpretation, following its use in the Asian-Pacific context. Furthermore, the findings of the current study will also serve to better inform the cultural and educational expectations of professional practitioners working in the field of comparative psychology. Closer to home, we are also a step nearer helping to better inform the parents of Asian infants who may wish to evaluate their children's relative (and comparative) developmental abilities, and thus increasing both the duration and depth of engagements they may seek and enjoy sharing with their own child(ren)'s early development, and/or in evaluating the putative effectiveness of any early infant training programs they may wish to attend.

References

- Bhat, G., Pardeshi, S., Kakrani, V., & Pratinidh, A. (2001). Making healthy baby contests more objective. *Indian Journal of Medical Science*, 55, 553-558.
- Curtin, R., Presser, S., & Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64(4), 413-428.
- Dickens, W. T., & Flynn, J. R. (2001). Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved. *Psychological Review*, 108(2), 346-369.
- Dickinson, A. R., & Zheng, E. (2012). *Ü-Baby Program Training Manual & User Guide*. Key Button International Education Inc. (MD, USA).

- Dickinson, A. R., Hui, J., & Zheng, E. (2011). *Course Primer in Preparation for the Certifications of Psychometric Test Administrator (CTA), & Test Consultant (CTC): ARL-ASQ-II*. Glück Psychological Services (HK, SAR).
- Kulkarni, M. L., & Mohan, G. L. (1999). Healthy Baby Contests-Beyond show, Beyond Objectivity Towards Health Education. *Indian Pediatrics*, 36, 417-418.
- Prajapathi, N. C. (1998). Healthy child contests: Making them more objective. *Indian Pediatrics*, 35, 1142.
- Squires, J., & Bricker, D. (2006) *Ages and Stages Questionnaires, Second Edition (ASQ-2™) A Parent-Completed Child-Monitoring System*. Chicago, USA: Paul H. Brookes Publishing Co.
- Yung, Y., & Dickinson, A. R. (2008). Being and Becoming Gifted: Enhancement effects of socio-intellectual study programs. *Proceedings of the 10th Asia Pacific Conference on Giftedness*. APCG, Singapore, 2008.

Received January 08, 2013

Revision received June 16, 2013

Accepted July 08, 2013