The feasibility of the Ages and Stages Questionnaire for the assessment of child development in a community setting in Nepal.

Merina Shrestha[1], Tor A Strand[2,3], Manjeswori Ulak[1], Ram K. Chandyo[1], Suman Ranjitkar[1], Mari Hysing[4], Laxman Shrestha[1], Ingrid Kvestad[4]

[1] Child Health Research Project, Department of Pediatrics, Tribhuvan University Teaching Hospital, Kathmandu, Nepal

[2] Department of Research, Innlandet Hospital Trust, 2629 Lillehammer, Norway

[3] Centre for International Health, University of Bergen, P.O. Box 78005020 Bergen, Norway

[4] Department of Psychosocial Science, University of Bergen, Bergen, Norway

[5] Regional Centre for Child and Youth Mental Health, Uni Research Health, Bergen, Norway

Correspondence: Dr Merina Shrestha, Assistant Professor, Department of Child Health, Tribhuvan University Teaching Hospital, Kathmandu, Nepal.

Email: drmerinashrestha@gmail.com

Acknowledgements

We would like express our gratitude to all the field staffs, children and families in Bhaktapur who participated in the study. Our thanks go to Prof. Prakash Sunder Shrestha for his encouragement and support. We are grateful to Child Health Research Project Team.

Funding
Thrasher Research Fund (Grant # 11512), and GC Rieber Funds

Abstract

Background: The Ages and Stages Questionnaire 3rd edition (ASQ-3) may be a feasible and cost-effective tool to screen children`s development in resource poor settings. We have assessed the feasibility of the ASQ-3 when conducted by fieldworkers in a community based nutritional interventional trial on early child development in Nepal.

Method**:** Six hundred children aged 6-11 months at risk of stunting were assessed by trained fieldworkers in their homes by the ASQ-3. Three fieldworkers performed standardization exercises and were double scored with a gold standard during the study period. Intra class correlations (ICC) were calculated to measure the inter-rater agreement. The internal consistency was expressed by standardized Cronbach`s alphas. The association between total ASQ score and preterm, low birth weight and stunted children is presented to give an estimate of the convergent validity.

Result: The mean total ASQ score for the 600 infants was 209.2±38.7. All mean scores in the current study, were consistently lower than in the American norm sample. The ICCs from the standardization exercises were initially good to excellent, but declined throughout the study period. The standardized alphas for the total score in the different age groups indicate good internal consistency, but varied in the subscales. Children who were preterm, children with low birth weight and stunted children scored substantially lower on the total ASQ score than those who were not.

Conclusion: Although the ASQ-3 “home” procedure seems to be feasible in Nepalese context, it still requires rigorous training to achieve acceptable inter-rater reliability and further adjustment to achieve satisfactory internal consistency. The cost of the tool is negligible and can be easily accessed. However, when “home” procedure is followed more keen observations and stringent training is required.

Key words: *Age and Stage Questionnaire, developmental screening, psychometric properties, LMIC*

Introduction

An estimated 250 million children under five years of age in low and middle income countries (LMIC) are at risk of not achieving their developmental potential (Black et al., 2017). Nepal is a poor country, ranked 145th out of 187 countries on the Human Development Index in 2015 (*Human Development Report*, 2015). Although the rate of mortality and morbidity in children under five in Nepal have declined, recent estimates demonstrate that about 42% of Nepalese children between 3 to 4 years of age have low scores on a childhood developmental index. Poverty, inadequate nutrition and stunting may add to the risk for developmental delay for these children (McCoy, 2016).

Early identification of developmental delay is a prerequisite for early targeted interventions, which help promote optimal developmental outcomes (Engle, 2011). For developmental assessment, comprehensive tests such as the Bayley Scale of Infant and Toddlers Development (Bayley) is often referred to as the gold standard. The Bayley scales provide broad information on a child's current development, and is administered directly with the child (Walder, 2009). Comprehensive tools LIKE the Bayley, however, are highly expensive and require administration by skilled professionals in order to achieve accurate scores. Thus, it is often not feasible to use such tools in resource poor settings like Nepal. Brief developmental assessment tools that are affordable and that can be administered by briefly trained medical personnel, are feasible and cost-effective alternatives to screen children for developmental delay in these settings.

The Age and Stages Questionnaire 3rd edition (ASQ-3) is an alternative developmental assessment instrument. ASQ-3 has emerged as a global development screening tool for children 2 to 66 months of age, in the general population (Singh, 2017; Small, 2018) and in children at risk, as for example premature babies (Schonhaut, 2013). When compared with the Developmental Assessment Scale for Indian Infants, the sensitivity and specificity of ASQ was found to be satisfactory in an Indian context (Juneja, 2012). ASQ-3 has been translated to several languages in Europe, Asia and South America, and have shown its adaptability in diverse cultural environments (Filgueiras, 2013; Small, 2018). However, there are many challenges in adapting the ASQ in different cultural settings, one of which is the insufficient rigor in following recommended steps of translation and adaptation (Clifford, 2017).

In order to evaluate the feasibility of a development assessment tool, the psychometric qualities such as the reliability and the validity need to be assessed within the given context (Fernald, 2009). Standardization exercises and double scorings are means to assess the reliability of a tool. In the literature, factors such as gestation, birth weight, growth and maternal level of education have been shown to be related to early child development (Grantham-McGregor, 2007), and are relevant measures to support the construct validity of the ASQ-3 when transferred to a new cultural setting.

In LMIC where educational levels and literacy rates are low, a limitation of using questionnaire-based screening instruments is the reliance on the parents' report. Hence, assessment methods like the ASQ "home procedure" might be an alternative. In the ASQ "home procedure", children are assessed by trained examiners with direct observation. The ASQ "home procedure" when applied by trained fieldworkers, have been found to be feasible in many previous studies (Handal, 2007; Kvestad, 2013).

Although there are a few studies on the ASQ in Nepal, the psychometric properties and the feasibility of the ASQ have not been thoroughly studied previously (Small, 2018). The aim of the current study is to assess the psychometric properties and the feasibility of the ASQ-3 "home procedure" in a Nepalese setting when carried out by trained fieldworkers in the home environment.

Methods

Study Sample

In the current study, children participated in a doubled blinded clinical trial entitled "The effect of Vitamin B12 supplementation in Nepali Infants on Growth and Development" registered under clinical trial number (CTClinicalTrials.gov: NCT02272842). The study was set in Bhaktapur, Nepal, and children aged 6 - 11 months at risk of stunting (length for age z-score < -1SD), whose family planned to reside in the area for the next 12 months, and whose parents consented to participate were included. Children with severe systemic illness requiring hospitalization, severe malnutrition (weight for height z-score < - 3SD) and with severe anemia (Hb < 7 g/dL) were excluded from the study. Children with ongoing acute infections such as fever or infection that required medical treatment, were temporarily excluded and enrolled after recovery.

Procedure

The study was conducted from April 2015 till February 2017. The children were identified by fieldworkers from immunization clinics or door to door home visits, and enrolled when the length was reconfirmed by a supervisor or physician at the field

office. A total of 600 children were enrolled and randomized to receive placebo or vitamin B12 supplements for 12 months. Enrollment procedures included collection of demographic information of the families, length and weight taking, blood sampling and developmental assessments. The developmental assessment was performed within 7 days of enrollment, preferable in the home of the child, but due to the destructions after major earthquakes in 2015 at the beginning of the study, some assessments were also conducted in open courtyards and temporary shelters. Assessments were conducted at the field office when it could not be performed at home.

ASQ-3

ASQ-3 is a screening tool to assess child development (Squires, 2009). It includes assessment of five developmental domains: communication, gross and fine motor, problem solving and personal social skills. For each subscale there are six questions, summing up to 30 questions for each age group. The ASQ-3 has 21 age specific questionnaires from 2 months till 66 months of age. The original ASQ was developed in the United States and the norms are based on a norming sample of 18,572 US children. The informants in the norm sample were mostly mothers, and most of the mothers had education level above high school.

In the current study the ASQ-3 “home procedure” which is described in the ASQ-3 manual and also used by others (Handal, 2007; Kvestad, 2015; Squires, 2009), was followed as mothers might not be able to understand and answer all the given questions. In this procedure, the items are answered based on direct observation by an examiner. We performed the assessments at the children`s homes, but the assessments could also be performed at the field office. Aided by a standardized kit of toys such as blocks, books and toys, trained fieldworkers administered the items with the child

through eliciting the relevant skills in the sessions. Tools and toys required for the assessment were prepared locally. For holding and banging we used inch sized wooden cubes, and for pincer grasp cheerios was used. Scoring was done on an opportunistic basis, meaning that the child did not need to perform serially as per the questionnaires. A few questions which were difficult to observe in the assessment setting, for example “making high-pitched squeals”, “repeating back sounds”, “putting foot in mouth while lying on floor”, “pushing arm through a sleeve”, “following simple commands without gestures”, “pointing to the wanted objects”, “lifting foot for shoes, socks or pant” were scored based on the report from the mother/caretaker. These items were marked as “not observed” in the data set.

Translation and cultural adaptions

For the study we used the questionnaires for 6, 8, 9, 10 and 12 months in order to cover all children in the age range of the study. Translation of the questionnaires was done following official recommendations for standards in translation and adaptations (Wild, 2005). The questionnaires were first translated from English to Nepali by the first author who have experience in the field of early child development. The Nepali translation was then back translated to English by a Nepalese Professor in English Literature. Finally, the back translation was reviewed by the first author and a Norwegian psychologist. The final version of the questionnaires was adapted after several discussions among the psychologists and fieldworkers. During the piloting, we found that the children did not like soft dolls, hence we replaced it with a plastic doll as Nepalese children are more familiar with these.

Training and Standardization

Three fieldworkers with educational level of at least grade 10, but with little

experience in developmental assessments were recruited for the study. A three days hands-on training on ASQ-3 was conducted. Ahead of the start of the enrollment, as a pilot we performed standardization exercises in 20 children. The first author who is a developmental pediatrician and experienced in developmental assessments in young children, served as the gold standard during standardization and the quality control. During the standardization procedure, one of the fieldworkers performed the assessment, while the other two fieldworkers and the person who served as the gold standard observed and scored independently. This method of standardization is inspired by the method that has previously been used in other studies, i.e. in intervention trials in India where early child development was an outcome (Taneja, 2005; Kvestad, 2013).

Throughout the study, 3% (20 children) of the assessments were double scored by the gold standard for quality assurance. During the double scoring procedure, the fieldworkers did the assessment in the presence of the gold standard. The gold standard and the fieldworkers scored independently and were blinded to each other`s scoring.

Ethics

The main study has obtained approval from the Nepal Health Research Council (NHRC, #233/2014) in Nepal and from the Regional Committee for Medical and Health Research Ethics (REC # 2014/1528) in Norway. Written informed consent from one of the parents (usually the mother) was obtained. For illiterate parents we took thumbprints in the presence of an impartial witness. Implementation of the study was as stated in the latest version of the Helsinki Declaration.

Statistical analyses

Data was coded according to the ASQ-3 manual; Yes = 10, Sometimes = 5 and Not Yet = 0. Following assessments, fieldworkers entered data into the iformbuilder data capture and management system (Zerion Software Inc.). The data was merged with the central server at the end of each day.

Data are presented as means (SD). We also present the means from the US sample from the ASQ-3 User Guide Manual (Jane Squires 2009). The Students t-test was used to compare the sample means. We calculated the intra-class coefficients to compare the fieldworkers' assessment with the gold standard during standardization and quality controls throughout the study. The ICCs were obtained from one-way random effects models for single measurements in Stata version 15 (Stata, College Station, TX) (Barnhart, 2007). Spearman`s correlation coefficient was calculated between the total ASQ-3 score and the five subscales across age groups, and standardized alphas was calculated for the total scores and the subscale scores for each age group (6, 8, 9, 10 and 12 months). We calculated the means (SD) according to gestation (preterm: < 37 weeks; term: $\geq$ 37 weeks), birth weight (<2500g, ≥2500g), maternal literacy (<5th grade, ≥5th grade) and stunting (<-2 length-for-age z-scores, ≥-1 length-for-age z-scores), and used a t-test to compare the means between the groups. All analyses were performed using STATA *version* 15.

Results

Demographic Characteristics

A total of 600 children were enrolled. The mean (SD) age was 8 (1.7) months and about 48% were female. About half of the children were first born, and lived in nuclear families. Only 10% were exclusively breastfed for 6 months. In about 35% of the families, both mother and father had an educational level up to grade 5. Most of

the mothers (62%) were housewives, and only about 5% of the fathers were not employed (Table 1).

ASQ sample distribution and comparison to US norms

Compared with the US sample, the Nepalese study children scored lower in all the subscales, with the largest differences in the problem solving and personal social subscales. The largest discrepancy is seen in personal social subscale across all the age groups, and a discrepancy is also seen in the gross motor subscale in higher age groups (Table 2).

Non-observable items

Most of the items could be scored directly during the observation. However, in the communication subscale, fieldworkers had to rely on caregivers' response 56% of the time on items like producing monosyllable or bisyllable sounds and repeating back. Regarding pointing to known objects when asked, fieldworkers could observe this for only 30% of the time. In the personal social subscale, items like grabbing foot while lying on the floor or putting the foot in the mouth and dressing by pushing arms through the sleeves and lifting the foot for shoes and socks, had to be asked in about 60% of the time, as these items could not be observed during the sessions.

Reliability measure

Inter rater agreement

Table 3 shows the ICCs between the gold standard and the fieldworkers. The ICCs during standardization, range from 0.57 (fine motor subscale) to 0.92 (gross motor subscale). During the study period the ICC values range from 0.36 (personal social subscale) to 0.63 (gross motor subscale).

Internal Consistency

Table 4 shows the Spearman correlation coefficients across all age levels between the total ASQ-3 score and the subscales, and between the different subscales. The correlations between the total ASQ-3 scores and the subscales are strong and the correlations between the five subscales are weak to moderate.

Table 5 shows the standardized Cronbach`s alphas for the total and subscale scores for each age level. For the total ASQ-3 score, the standardized alphas range from 0.60 (8 months) to 0.77 (6 months), all indicating satisfactory internal consistency. For the subscales, the standardized alphas range from 0.2 (personal social domain at 9 and 12 months) to 0.7 (gross fine motor domain at 12 months). In the fine motor subscale, there was a negative average covariance for 3 of the age groups (9, 10 and 12 months).

Construct validity

Table 6 shows the mean differences in total ASQ-3 score according to gestation, birth weight, maternal literacy and growth. Children who were preterm, who had low birth weight and stunted children had lower scores compared to those who were not. There was no significant difference in total ASQ score between children of mothers with educational level up to grade 5 compared to children with more than 5 years of education. However, in linear regression models, we found that for every change in the level of education, there was on an average a 2.99 [(95% CI 0.69, 5.3), p=0.011] point change in the ASQ-3 score. Similarly, for each unit change in Length for age (LAZ), there was 9.8 [(95% CI 4.5,15.2), p = 0.001] point change in ASQ score.

Discussion

The current study assessed the feasibility and psychometric properties of the ASQ-3 "home procedure" when conducted by trained fieldworkers in 600 Nepalese children. Our results show that the mean scores in the Nepalese children were consistently lower than the US normative sample. Inter rater agreement ranged from fair to excellent during standardization, and from poor to good in the double scorings during the study. This suggests that while the training was successful, there was a drift in the fieldworkers scoring throughout the study. For the total ASQ-3 score, results indicate satisfactory internal consistency, while the alphas vary between the subscales and the different age ranges. The results demonstrate that the ASQ-3 discriminate between children known to be at risk of poor development, such as children with LBW, who are preterm, stunted and where the mothers have low educational levels, giving support to the construct validity of the tool in a Nepalese setting.

All ASQ subscale scores were lower compared to the US normative sample in the present study. Scores are also lower than previously reported mean levels in Norwegian children at 6 months (Alvik, 2011), and children at 12 months in South Korea and South Africa (van Heerden, 2017). The comparable low scores in the current study might be due to a number of factors. All included infants had a length-for-age z-score less than -1 and were at risk of stunting (Handal, 2007). Stunting in children is a well-established risk factor for adverse early child development which may be part of the explanation for the lower mean scores in the current study compared to the US sample (Walker, 2011). Another reason for the low scores in the current study might be the lack of awareness amongst Nepalese mothers both on the milestones in early child development, and in the different activities that can be provided to the children at home (Shrestha, 2017). The largest discrepancies in scores

were in the personal social scale, which may indicate that children are given less opportunity to explore, self-eat and dress at very young age in the Nepalese setting that in western countries. Differences between children in LIMC and the norms on the personal social subscale are in accordance with results in previous studies (Gladstone, 2008).

The evaluation of the internal consistency of the ASQ-3 when transferred to a Nepalese study setting was done by the Spearman correlations and the standardized alphas. The strong correlations coefficients between the total score and the subscale scores support the consistency of the scale. The comparable lower scores between the subscale scores have also been found by others (Kvestad, 2013), and could demonstrate that although there are some overlap between the scores on the different subscales, these are measuring different domains. The standardized alphas for the total score in the different age groups indicate good internal consistency when the scale is used on a total score. Our findings are similar to the results when the Georgian ASQ-3 was applied to children 1-66 months of age where alpha values for all age groups varied from 0.643 to 0.824 across areas (Zirakashvili, 2018). However, for the subscales, the alpha values in the present study were highly variable with several very poor values, and even a negative average covariance for three age groups in the fine motor subscale. These values indicate that there are items in the translated Nepalese version that are inconsistent with the other items in the subscale, and that these items may not be suitable in assessing the relevant developmental domain in this cultural setting. These inconsistencies may have various reasons, such as difficulties in the cultural adaptations of the items and to difficulties in the scoring of the observed behavior. Further investigation of these items is necessary with the aim of additional adjustments to improve the internal consistency of the scales. Our findings

are similar to other studies (Kvestad, 2013), and call for new adaptions and piloting in order to increase the usability of the questionnaires in a Nepalese setting. Until further adjustments are done, our results indicate that the ASQ-3 in this setting for now is most reliable on the total score, and that the subscale scores should be interpreted with caution.

In our study, we found that the total ASQ-3 score discriminated well in premature children, LBW children, and in children who were stunted. Also, there was a linear relationship between the ASQ-3 scores and maternal level of education. ASQ has previously been found to be an effective screening tool in identifying developmental delay when applied on premature babies (Richter, 2007). The construct validity of the tool has further support by the linear relationship between the ASQ scores and height for age z-scores in a North Indian setting (Kvestad, 2015). It has also been demonstrated that the ASQ can identify more children with questionable development or delays than the Pediatric Developmental Impression with increased referral rates at 12 months of age (Hix-Small, 2007). Taken together our results with that of previous studies give support to the construct validity of the ASQ-3 in this Nepalese setting.

We have used the ASQ "home procedure" as outlined in the ASQ-manual and as used by others (Handal, 2007; Kvestad, 2015; Squires, 2009). This procedure requires training and standardization to ensure reliable data. The ICCs for the standardization exercises were fair to excellent, indicating that the training was successful in establishing reliable assessments across raters. However, the ICCs on the quality checks during the study were poorer indicating measurement drift throughout the study period. Our result could indicate that a more thorough and repeated training would be required in order to reach higher agreement between the testers and prevent

drift. The previous study in India, also used the ASQ-3 “home procedure” (Kvestad, 2015). However, this study demonstrated consistently strong ICCs throughout the study. The fact that in the Indian study all assessments were carried out in the same room at the study clinic, could have given a context to the assessments that to a higher degree prevented measurements drift than in Nepal. In Nepal, all assessments were done in the child`s home setting, representing an extra challenge for the examiners to do standardized assessments. Furthermore, to use fieldworkers with more experience in the assessment of early child development could be necessary. Finally, close monitoring of the assessment process is a prerequisite in large studies when using fieldworkers with little previous knowledge and experience in assessing early child development to prevent examiners drift.

Although, the ASQ “home procedure” use the direct observation of the child`s performance, two of the subscales (communication and personal social) required that most of the items were scored based on caregivers’ report and was accordingly more subjective. In the communication subscales, fieldworkers had to rely on caregivers’ response in 60% of the items as children often did not vocalize during the assessments. In a previous study, we show that most of the Nepalese mothers think that they should begin to talk to their children at age of 11.5 ± 8.6 months, and very few mothers knew when to start shared book reading (Shrestha, 2017). The same study also demonstrates that there is a lack of knowledge of early child development in this population, which could also influence the reliability of the answers of these mothers.

A strength of the study is the rigorous translation and piloting of the ASQ-3 including forward and backward translation, involvement of local pediatricians and

fieldworkers and piloting in the field before enrolling children (Clifford, 2017). Standardization of fieldworkers along with double scoring with gold standard during the study period also added to its strength. Due to capacity limitations we only have double scorings in 3 % of the cases (20 children) for the quality control throughout the study period. This number is low, which represent a limitation to the study. Although the ASQ "home procedure" is doable in this setting, the ICCs indicate that there are differences in the scoring among the fieldworkers. To calculate the ICCs at regular intervals throughout the study, rather than only at the end of the study, could have reinforced additional training exercises and improved the agreement between the fieldworkers. The generalization of the results is also limited by the fact that this is a high-risk sample in the small community of Bhaktapur.

Conclusion

In a Nepalese context, although the ASQ-3 system seems to fulfill some of the criteria of feasibility like low cost, easily accessible and manageable training as proposed by the expert group (Fischer, 2014), when "home procedure" is followed, the inter rater agreement during the study was poor. Thus, more keen observations and stringent training is required. Moreover, our result demonstrates that in the current version of the ASQ-3 translated to Nepalese, the total score should be considered rather than single subscales for developmental assessments.

Key messages

- The Ages and Stages Questionnaire 3rd edition (ASQ-3) may be a feasible and cost-effective tool to screen children`s development in resource poor settings.
- The ASQ-3 Nepalese "home procedure" undertaken by fieldworkers requires keen observations and stringent training.
- ASQ-3 discriminates between children with various risks of poor development, based on gestation length, birth weight, linear growth, and maternal educational level, giving support to the construct validity of the tool in a Nepalese setting.

References

Alvik, A., & Grøholt, B. (2011). Examination of the cut-off scores determined by the Ages and Stages Questionnaire in a population-based sample of 6 month-old Norwegian infants. *BMC pediatrics, 11*(1), 117.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., . . . Shiffman, J. (2017). Early childhood development coming of age: science through the life course. *The Lancet, 389*(10064), 77-90.

Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of biopharmaceutical statistics, 17*(4), 529-569.

Clifford, J., Squires, J. K., & Murphy, K. (2017). Not Lost in Translation: Modifying the Ages & Stages QuestionnairesTM for Use in Cross-Cultural Contexts. *Current Developmental Disorders Reports, 4*(4), 130-136.

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., . . . Ertem, I. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet, 378*(9799), 1339-1353.

Fernald, L. C., Kariger, P., Engle, P., & Raikes, A. (2009). Examining early child development in low-income countries: a toolkit for the assessment of children in the first five years of life: World Bank.

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.

Fischer, V. J., Morris, J., & Martines, J. (2014). Developmental screening tools: feasibility of use at primary healthcare level in low-and middle-income settings. *Journal of health, population, and nutrition, 32*(2), 314.

Gladstone, M., Lancaster, G. A., Jones, A., Maleta, K., Mtitimila, E., Ashorn, P., & Smyth, R. (2008). Can Western developmental screening tools be modified for

use in a rural Malawian setting? *Archives of Disease in Childhood, 93*(1), 23-29.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., & Group, I. C. D. S. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet, 369*(9555), 60-70.

Handal, A. J., Lozoff, B., Breilh, J., & Harlow, S. D. (2007). Sociodemographic and nutritional correlates of neurobehavioral development: a study of young children in a rural region of Ecuador. *Revista Panamericana de Salud Publica, 21*, 292-300.

Heo, K. H., Squires, J., & Yovanoff, P. (2008). Cross‐cultural adaptation of a pre‐school screening instrument: comparison of Korean and US populations. *Journal of Intellectual Disability Research, 52*(3), 195-206.

Hix-Small, H., Marks, K., Squires, J., & Nickel, R. (2007). Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. *Pediatrics, 120*(2), 381-389.

Human Development Report. (2015). Retrieved from http://hdr.undp.org/sites/default/files/2015_human_development_report.pdf

Janson, H. (2003). Influences on participation rate in a national Norwegian child development screening questionnaire study. *Acta Pædiatrica, 92*(1), 91-96.

Juneja, M., Mohanty, M., Jain, R., & Ramji, S. (2012). Ages and Stages Questionnaire as a screening tool for developmental delay in Indian children. *Indian pediatrics, 49*(6), 457-461.

Kvestad, I., Taneja, S., Hysing, M., Kumar, T., Bhandari, N., & Strand, T. A. (2015). Diarrhea, stimulation and growth predict neurodevelopment in young north Indian children. *PloS one, 10*(3), e0121743.

Kvestad, I., Taneja, S., Kumar, T., Bhandari, N., Strand, T. A., & Hysing, M. (2013). The assessment of developmental status using the Ages and Stages questionnaire-3 in nutritional research in north Indian young children. *Nutrition journal, 12*(1), 50.

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., . . . Fink, G. (2016). Early childhood developmental status in low-and middle-income countries: national, regional, and global prevalence estimates using predictive modeling. *PLoS Medicine, 13*(6), e1002034.

Richter, J., & Janson, H. (2007). A validation study of the Norwegian version of the Ages and Stages Questionnaires. *Acta Paediatrica, 96*(5), 748-752.

Schonhaut, L., Armijo, I., Schönstedt, M., Alvarez, J., & Cordero, M. (2013). Validity of the ages and stages questionnaires in term and preterm infants. *Pediatrics*, peds. 2012-3313.

Shrestha, M., Ulak, M., Strand, T. A., Kvestad, I., & Hysing, M. (2017). How much do Nepalese mothers know about child development? *Early Child Development and Care*, 1-8.

Singh, A., Yeh, C. J., & Blanchard, S. B. (2017). Ages and stages questionnaire: a global screening scale. *Boletín Médico Del Hospital Infantil de México (English Edition), 74*(1), 5-12.

Small, J. W., Hix-Small, H., Vargas-Baron, E., & Marks, K. P. (2018). Comparative use of the Ages and Stages Questionnaires in low-and middle-income countries. *Developmental Medicine & Child Neurology*.

Squires, J., & Bricker, D. (2009). Ages & stages questionnaires, (ASQ-3). *A parent-completed child monitoring system*.

Taneja, S., Bhandari, N., Bahl, R., & Bhan, M. K. (2005). Impact of zinc

supplementation on mental and psychomotor scores of children aged 12 to 18 months: a randomized, double-blind trial. *The Journal of pediatrics, 146*(4), 506-511.

van Heerden, A., Hsiao, C., Matafwali, B., Louw, J., & Richter, L. (2017). Support for the feasibility of the ages and stages questionnaire as a developmental screening tool: a cross-sectional study of South African and Zambian children aged 2-60 months. *BMC pediatrics, 17*(1), 55.

Walder, D. J., Sherman, J. C., & Pulsifer, M. B. (2009). Neurodevelopmental Assessment. *Evidence-Based Practice in Infant and Early Childhood Psychology*, 167-205.

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., . . . Lozoff, B. (2011). Inequality in early childhood: risk and protective factors for early child development. *The Lancet, 378*(9799), 1325-1338.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee‐Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient‐reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in health, 8*(2), 94-104.

Zirakashvili, M., Gabunia, M., Tatishvili, N., Ediberidze, T., Lomidze, G., Chachava, T., & Hix-Small, H. (2018). Cultural Adaptation and Psychometric Validation of the Ages and Stages Questionnaires for use in Georgia. *Journal of Child and Family Studies, 27*(3), 739-749.

Table 1: Demographic characteristics of 600 Nepalese infants

Child	
Mean age of child in months (SD[1])	8 (1.7)
Female child (%)	291 (48.5)
Birth order (%)	
1st	292 (48.7)
2nd	229 (38.1)
3rd	79 (13.2)
Mean birth weight in grams (SD)	3041 (1418)
Exclusive breastfeeding for 6 months or more	64 (10.6)
Family type	
Nuclear	308 (51.3)
Joint	292 (48.7)
Mother	
Mean age (SD)	27 (4)
Literacy	
Illiterate up to grade 5	211 (35.5)
Grade 5 to 10	130 (21.8)
Grade 11 and 12	148 (24.8)
Bachelor	73 (12.2)
Master or above	33 (5.5)
Occupation	
No working mother/Agriculture	371 (61.8)
Carpet worker	17 (2.8)
Daily wage earner	73 (12.2)
Services	73 (12.2)
Father	
Mean age (SD)	25 (8)
Literacy	
Illiterate up to grade 5	221 (37)
Grade 5 to 10	113 (19)
SLC to grade 12	148 (24.8)
Bachelor	88 (14.7)
Master or above	26 (4.3)
Occupation	
No working/Agriculture	32 (5.3)
Carpet worker	7 (1.2)
Daily wage earner	233 (39)
Services	113 (19)
Self employed	173 (29)

[1]SD= Standard Deviation

Table 2: Participants mean (SD[1]) scores in the 600 Nepalese infants compared with the US[2] norm sample mean (SI

Month	N[3]	Communication			Gross Motor			Fine motor			Problem Solving			Personal Socia	
		Mean	SD	mean diff.	Mean	SD	mean diff.	Mean	SD	mean diff.	Mean	SD	mean diff.	Mean	SD
	499														
6	**168**														
Nepal		43.9	10.3	-5	40.4	12.3	-4.1	46.1	15	-2.8	39.4	12.8	-11	32.9	13.1
US		48.9	9.6		45.6	11.6		48.9	12		50.4	11.3		48.3	11.5
8	**161**														
Nepal		44.5	7.4	-7.9	46.6	10	-5.4	50.2	8.6	-5.5	43.8	11.6	-10	43.7	9.7
US		52.4	9.6		52	10.7		55.7	7.8		53.9	8.8		53.3	8.7
10	**84**														
Nepal		39.8	10.5	-8.4	39	11.9	-14	46.9	6.6	-7.8	47.2	11	-4.9	33.3	10.7
US		48.1	12.6		53	11.4		54.7	8.3		52.1	9.8		49.4	11.1
12	**86**														
Nepal		42.5	9.6	-0.7	38.3	16	-11	46.1	8.2	-6.1	44.3	11.6	-4.6	35.9	12.5
US		43.2	13.7		49.9	14.2		52.2	8.8		48.9	10.8		45.7	12

[1]SD= Standard Deviation, [2]US=United States, [3] US norm sample do not include 9 months questionnaire, hence

101 children from the study could not be compared with the US sample norms

Table 3. Intraclass correlations (ICC[1]) between the field workers and a gold standard during standardization and quality control

	Standardization Exercise	Quality Control
Total	0.88	0.52
Communication	0.91	0.54
Gross motor	0.92	0.53
Fine motor	0.57	0.63
Problem solving	0.87	0.47
Personal social	0.71	0.36

[1]ICC inter-rater agreement measures: Poor: Less than 0.4; Fair : Between 0.40 and 0.59; Good: 0.60 and 0.74 ; Excellent: 0.75 and 1.00

Table 4: Spearman rank correlation coefficient between ASQ-3[1] total score and subscales scores and between the subscales

	Total	Communication	Gross motor	Fine motor	Problem solving	Personal social
ASQ-3						
Total	1					
Communication	0.52	1				
Gross motor	0.73	0.21	1			
Fine motor	0.66	0.28	0.33	1		
Problem solving	0.64	0.14	0.38	0.40	1	
Personal social	0.63	0.25	0.33	0.30	0.17	1

[1]ASQ-3 = Ages and Stages Questionnaire, 3rd edition

Table 5: Standardized Cronbach`s Alphas[1] for the total ASQ-3 score and the subscale scores for each age ra

	N	Total	Communication	Gross motor	Fine motor	Problem Solving	Personal Soci
6months	168	0.77	0.29	0.44	0.57	0.49	0.18
8months	161	0.60	0.01	0.50	0.46	0.27	*
9months	101	0.70	0.19	0.45	*	0.37	0.19
10 months	84	0.60	0.10	0.35	*	0.34	0.32
12 months	86	0.64	0.16	0.65	*	0.27	0.20

[1]Alpha (α) >0.80 - highly, α = 0.60-0.80 – satisfactory, and α = 0.40-0.60 – moderate internally consistent, * - negative average covariance.

Table 6: Total ASQ-3 scores according to gestation, birth weight, maternal literacy and stunting and the mean difference (95% between the subgroups.

	N=600 (%)	Total ASQ-3 score	Mean Difference (95% CI)
Gestation			
Preterm	62(10.3)	193	18 (7.9, 28.2) ***
Term	538 (89.6)	211	
Birth weight			
<2500gm	115(19.2)	195	17.6 (9.8, 25.4) ***
≥ 2500gm	485(80.8)	212.6	
Maternal literacy			
<5th grade	22(37.2)	205.7	5.7 (0.7, 12.1)
≥ 5th grade	377(62.8)	211.4	
Stunted			
Yes	194 (32.5)	203.2	9.1 (2.4, 15.7) ***
No	403 (67.5)	212.3	