



## **Grow Along With Me Evaluation**

### **Reliability Audit**

Prepared by: Chris Jensen-Ross, Brittany Thompson & Dr. David Gregory

Prepared for: the Grow Along with Me Steering Committee, Jan. 20, 2010

## Grow Along With Me Evaluation: Executive Summary

### Ages and Stages Questionnaire: Validity and Reliability

2008-09

Purpose: The purpose of this study was to examine Ages and Stages Questionnaire (ASQ) data reliability regarding scoring, data entry, and the internal consistency of the instrument.

Sample and methodology: A random sample of 300 charts was selected for audit. Data were examined for completeness and accuracy.

*Cohen's Kappa* was used to determine reliability between scores:

- a) documented by the person completing the ASQ in comparison to scores calculated by a consistent reviewer;
- b) available on the questionnaire and those entered on the summary sheet; and,
- c) between written and computerized data (Master Database).

*Internal consistency* of the Ages and Stages Questionnaire (Version 2) when used with this Southern population was examined.

#### Results:

300 charts were audited. Results are generalizable to the entire population at a 95% confidence level (+/-5.1).

A Cohen's Kappa error rate of 5.4% was found for domain-specific score calculations (e.g. fine motor, gross motor) (in comparison to an acceptable error rate of 5.0%). Incorrect transfer of data between the ASQ, ASQ summary sheet and/or the Master Database occurred in 5.7% of the questionnaires audited. The occurrence of incorrect transfers between the ASQ and summary sheet, and the summary sheet and computerized data was 3.5%. The most common source of

data error (9%) was the use of an incorrect age-specific ASQ. Similarly, over 6% of ASQs completed for premature infants were not “age-adjusted”.

Audits of internal consistency showed high correlation (100%, n=15/15,  $p < .001$ , 95% confidence) between domain scores and overall ASQ scores. Internal consistency between age-specific questions and domain scores fell below acceptable standards ( $\alpha > .70$ ) 57% of the time (n=23/40).

## Recommendations

Based on the audit findings, it is recommended that:

- 1) The Grow Along with Me Steering Committee:
  - a) Provide ongoing education regarding ASQ administration, scoring, data collection and follow-up process;
  - i) Highlight the correct response to developmental areas where the child previously performed, but no longer does due to the acquisition of more advanced skills (e.g. crawling replaced by walking). Squires et. al (1999) state that “If parents mistakenly answer *not yet* or *sometimes* to an easier item... but *yes* to a more advanced item..., the score for the earlier item should be changed to 10 (for a *yes* response) before computing the total area score;” (p. 66)
  - b) Promote consistent use of developmental screening tools across all service providers;
  - c) Reinforce the calculation of ratio-based scores with staff responsible for scoring questionnaires;
  - d) Stress the importance of Corrected Date of Birth (CDOB) when selecting age-specific questionnaires for premature infants;

- e) Promote the collection and data entry of question level data in addition to summary data (including summary sheet, information sheet, and questionnaire). This would be particularly useful for future validity and reliability audits;
  - f) Encourage computerization of Ages and Stages data to reduce data transfer error and calculation error rates; and
  - g) Consider use of ASQ Version 3 in relation to identified issues that have been addressed in the newer version of the tool (e.g. consistency of wording across questionnaires, corrected date of birth).
- 2) ASQ developers should:
- a) Examine ASQ rigor and generalizability when used with a non-American, culturally diverse population.

## Grow Along With Me Evaluation. Reliability Audit

### *Introduction*

Confidence that data are valid and reliable is foundational to evidence-informed decision-making. Data rigor must be substantiated for the healthcare provider to have confidence in the data (Grover, Hammermeister and Shroyer, 1995). The purpose of this audit was to determine the reliability and internal consistency of the Ages and Stages Questionnaire in conjunction with the Grow Along with Me Project, Alberta Health Services Southwest Zone (former Chinook Health Region).

### *Background*

Information regarding the growth and development milestones achieved by children aged 18-36 months (who participated in the Grow Along with Me Project [GAWM, 2007-09]) was obtained through the Ages and Stages Questionnaire (ASQ, Version 2). Eight age-specific ASQs (18, 20, 22, 24, 27, 30, 33 and 36 months of age) are available. Each questionnaire contains 30 questions which are divided into the following five areas of development:

- Communication;
- Gross motor;
- Fine motor;
- Problem-solving; and,
- Personal-social development.

Parents are invited to outline any general concerns in an “overall section” at the end of the questionnaire.

ASQs were completed by parents/guardians at a Well Child Clinic check-up or in conjunction with a visit to a GAWM partner agency<sup>1</sup>. Trained staff (guided by explicit coding instructions [Squires, Potter & Bricker, 1999, p. 73]), converted the responses to color-coded scoring sheets. Scores were used to determine the child's progress in each developmental area. Where appropriate or at parent request, referrals for in-depth evaluation were initiated<sup>2</sup>. Completed ASQs, scoring sheets and research consent forms were then forwarded to Children's Community Assessment, Rehabilitation and Education (C.A.R.E.) Services where the scores were entered into the C.A.R.E. Master Spreadsheet.

The validity of the ASQ (version 2) has been previously established (Squires et al., 1999) and contributed to the selection of the ASQ as the tool of choice for GAWM, an Alberta Innovation Project. The performance of 1,763 children using the ASQ was compared with their performance using established standardized developmental tests. Squires et al. established that overall agreement across questionnaires was 83% (range 76% - 91%)<sup>3</sup>. Tests of reliability (e.g. inter-rate reliability and test-retest) exceeded 90%.

The questions of interest for this audit focused on reliability of the ASQ when used with this Southern Alberta population. Specifically, the questions included:

---

<sup>1</sup> Partner agencies included: Families First, Parent Link Centres, Family Centre, Family & Community Care Services (FCSS)

<sup>2</sup> Cutoff points have been determined for each area of development on each questionnaire based upon mean and spread (2 standard deviations from the mean) based upon 7,900 questionnaires (Bricker, Squires & Potter, 1999, p. 73).

<sup>3</sup> Underreferral rates ranged from 1% to 13%) while overscreening rates ranged across intervals from 7-16%.

1. How reliable is the:
  - a. Ages and Stages Questionnaire (ASQ) scoring (in comparison to scoring guidelines, Squires et al., 1999); and,
  - b. Data in the existing Grow Along with Me (GAWM) Master Database spreadsheet (in comparison to the calculated score)?
2. What is the internal consistency of the Ages and Stages questionnaire?

### ***Methodology***

Each ASQ received by Children's C.A.R.E. services was randomly assigned a number. Questionnaires were used if ASQ (Version 2) was utilized, if the summary sheet **and** questionnaire were available, and if the ASQ was received by Children's Care in 2008-2009; questionnaires not fitting this criteria were excluded. For ASQs completed by Children's C.A.R.E. staff, summary scores were available in electronic form (Meditech database). The sample size required to generalize audit findings to the projected population (e.g. ASQs received in 2008-09, 95% confidence  $\pm 5$ ) was determined. Consequently 300 questionnaires were randomly selected for audit using Statistical Package for the Social Sciences (SPSS).

The audit was completed in two phases. In Phase 1, 100 ASQs were audited. To select the ASQs for the first audit, the assigned numbers of the 300 selected ASQs were listed in random order. The first ASQ to be audited was randomly selected from the list and every 5<sup>th</sup> ASQ was chosen for the audit. In Phase 2, the remaining 200 questionnaires were audited.

Reliability analysis was completed using Microsoft Excel and PAWS Statistical Analysis Software. Cohen's Kappa was used to establish reliability between expected and actual ASQ

domain scores, domain summary scores and Master Spreadsheet values. Kendall's Tau-b<sup>4</sup> and Cronbach's coefficient alpha were used to establish internal consistency (e.g. between domain questions, domain summary scores and overall summary scores).

The audit code book and definitions can be found in Appendixes A and B.

## ***Results and Discussion***

### **Kappa-Alpha Reliability**

In total, 300 ASQs were audited. With a total of 1607 ASQs received by Children's C.A.R.E. in 2008-09<sup>5</sup>, this provided the ability to generalize the findings of this audit to the study population with 95% confidence (+-5.1%). The Kappa-Alpha reliability percentage of error was 5.7% (17/300). This included the total number ASQ questionnaires that had incorrect transfers of data between questionnaire, summary sheet, and computerized database. When examined in greater detail,

- The proportion of error in the transfers of data between the questionnaire and the summary sheet was 4.7% (14/300) (Kappa-Alpha reliability error);
- The proportion of error in transfer of data between the summary sheet and the Master Database was 2.3% (7/300); and,
- The proportion of error in the transfer of data between the questionnaire and summary sheet, and the summary sheet and master database was 3.5% (21/600).

---

<sup>4</sup> Kendall's tau-B was selected (rather than Pearson's correlation) due to the non-parametric distribution of the ordinal data

<sup>5</sup> Duplicate age-specific ASQs excluded

- The total proportion of incorrect domain-specific calculations or data transfers was 5.4% (81/1500).

According to Beaman et al. (2004), the acceptable percentage of error for surveys and quantitative research (with a sample of at least 100) was 5%. Based on Kappa alpha reliability estimates, the percentage of error between the ASQs and the Master Database was within an acceptable percentage of error (5%).

Areas of concern included the proportion of incorrectly selected age-specific ASQs (9%, 27/300) and incorrect adjustment for prematurity<sup>6</sup> (6.3%, 19/300). Error rates are not within the acceptable range and threaten the validity of the ASQ findings for these children.

**Table 1: Validity & Reliability: ASQ 2008-09 (n=300)**

<b>Area of Inquiry</b>	<b>Percentage (/100)</b>
Questionnaires excluded due to differing edition of ASQ (from 2 <sup>nd</sup> edition)	0%
Total number of Questionnaires with incorrect transfers of data within the Kappa-Alpha Reliability table (A vs. B, B vs. C)	5.7% (n=17/300)
Total incorrect transfers of data between A and B (Questionnaire and Summary Sheet) in Kappa-Alpha Reliability table	4.7% (n=14/300)
Total incorrect transfers of data between B and C (Summary Sheet and Master Database) in Kappa-Alpha Reliability table	2.3% (n= 7/300)
Total incorrect transfer between A and B and B and C (up to two transfers per questionnaire)	3.5% (n=21/600)
Total Questionnaires with the incorrect month of ASQ for the child's age	9% (n=27/300)
Incorrect calculations of a child's date of birth. Due to premature birth and Corrected Date of Birth (CDOB) not being used or calculated correctly	6.3% (n=19/300)
Total number of incorrect calculations within each domain (5 domains/Questionnaire)	5.4% (n=81/1500)

<sup>6</sup> Corrected Date of birth is to be used for children born more than 3 weeks prematurely up the age of two years.

## Kendall's Tau-B Correlational Analysis

In total 299 charts were included in the Kendall's Tau-B correlational analysis.<sup>7</sup> Correlation between developmental area scores and overall scores are presented in the following table:

Table 2 Correlation between developmental area and overall scores by age-specific ASQ

Correlations between area and overall score on the (8) age-specific questionnaires ^Kendall's tau_b (N=299)						
Age interval (in months)	n	Communication	Gross motor	Fine motor	Problem-solving	Personal-social
18	147	.61**	.31**	.49**	.58**	.43**
20	53	.58**	.46**	.65**	.65**	.58**
22	15	.47*	.46	.34	.71**	.70**
24	21	.62**	.36	.47*	.20	.39
27	10	.82**	.61**	.58	.72**	.65**
30	17	.62**	.44	.60*	.54**	.48*
33	7-8	.57	.50	.23	.46	.48
36	26	.66**	.37	.52**	.63**	.60**

\* Correlations are significant at  $p < .05$

\*\* Correlations are significant at  $p < .001$

With the exception of the 33-month questionnaire (where no significant correlations were detected), significant relationships between the area and overall scores were found 77.1% ( $n=27/35$ ) of the time.

Consistent with the Squires et al. (1999), correlation analysis showed significant relationships for all developmental areas and overall score for the total population (see Table 3).

Table 3 Correlation between developmental area and overall scores (collapsed across questionnaires [all ages])

Kendall's tau_b correlation between area scores collapsing across questionnaires					
Area	Communication	Gross motor	Fine motor	Problem solving	Personal-social
Communication					
Gross motor	.15*				
Fine motor	.23**	.22**			
Problem solving	.33**	.22**	.34**		
Personal-social	.29**	.16*	.29**	.34**	
Overall	.62**	.37**	.49**	.60**	.50**

Note: \*\* Correlations are significant at  $p < .001$  \* $p < .005$  (2-tailed)

In contrast to Squires et al. (1999), internal consistency was calculated using Kendall's Tau-B (rather than Pearson's correlation) because of asymmetrical data distribution. For the combined

<sup>7</sup> One chart was excluded due to question scores exceeding upper-limit criteria.

ASQs (all age-specific questionnaires), developmental domains and overall scores (the sum of domain-specific scores) were significantly related. This would seem to indicate a high degree of internal consistency across domains.

While all correlations were found to be significant at  $p < .001$  in the original tool, significance levels were lower for the GAWM sample. Age-specific cross-domain correlational ranges for the GAWM sample are compared to the ASQ version 2 correlation ranges (Squires, Bricker & Potter, 1999) in the table below:

**Table 4 Comparison of GAWM project and ASQ reference value correlation ranges (age-specific questionnaire overall scores and developmental area scores (all domains))**

<b>Correlation Range</b>		
<b>Age-Specific Questionnaire</b>	<b>GAWM Project</b>	<b>ASQ Reference Values</b>
20	.46 - .65	.66 - .77
24	.20 - .62	.63 - .76
30	.44 - .62	.69 - .83
36	.37 - .66	.77 - .83

Source: Squires, Bricker & Potter, 1999, p. 147

Lower correlations found in this project in comparison to the original study may be related to smaller sample size or may reflect different characteristics of this Canadian population in comparison to their American-based counterparts.

## Cronbach's Alpha

**Table 5. Cronbach's coefficient alpha. Standardized alphas by developmental area and age interval**

<b>Cronbach's coefficient alpha. Standardized alphas by developmental area questions and age interval</b>					
<b>interval (in mon)</b>	<b>Communication</b>	<b>Gross motor</b>	<b>Fine motor</b>	<b>Problem-solving</b>	<b>Personal-social</b>
18	.69	.63	.54	.57	.47
20	.76	.86	.77	.63	.83
22	.73	.84	.54	.71	.66
24	.81	.26	.35	-.03	-.16
27	.77	.81	.73	.40	.29
30	.85	.04	.68	.73	.25
33	.60	.27	.59	.56	.69
36	.73	.02	.83	.71	.32

Six questions make up each developmental area score. For the communication area, alphas ranged from .60 at 33 months to .85 at 30 months. For the gross motor area, alphas ranged from

.02 at 36 months to .86 at 20 months. For the fine motor area, alphas ranged from .35 at 24 months to .77 at 20 months. For the problem-solving area, alphas ranged from -.03 at 24 months to .71 at 22 and 36 months. For the personal-social area, alphas ranged from -.16 at 24 months to .83 at 20 months (see Table 3). The proportion of acceptable scores by age-specific questionnaire were: 80% (4/5) 20 months, 60% (3/5) 22, 27 and 36 months, 20% (1/5) 24 and 33 months, and 0% (0/5) 18 months.

Morgan, Leech, Gloechner & Barrett (2007) stated, “Alphas should be positive and usually greater than .70 in order to provide good support for internal consistency reliability.” (p. 129) Over half (57.5%, 23/40) of age and domain specific internal consistency ratings fell below acceptable levels; Negative alpha scores (5%, 2/40) were reported for two domains in the 24-month questionnaire (problem-solving and personal social). With the exception of communication, more than half (62.5%, =>5/8) of age-specific domain alphas fell below acceptable reliability levels<sup>8</sup>.

While internal consistency between-domains is high, within-domain ratings are low – possibly reflecting the need for domain-specific question refinement. Low internal consistency ratings for 18-24 month questionnaires found in this population sample persist for fine motor, problem-solving and personal social domains in ASQ (Version 3).

### ***Limitations***

- While the ASQ has been published in several different languages, cultural and ethnic variables are not captured through the ASQ;

---

<sup>8</sup> The proportion of acceptable scores by developmental area was: 75% (6/8) communication, 37.5% (3/8) gross motor, fine motor and problem-solving and 25% (2/8) personal-social.

- Several ASQ questionnaires were excluded because they could not be located in the GAWM master spreadsheet database. This may have been a consequence of duplicate questionnaires or filing errors (wrong spelling of the child's name). ASQ questionnaire postal codes may or may not reflect the child's residence as they are entered by the person completing the questionnaire.
- Incomplete or missing information about the child's actual, versus expected date of birth is critical, particularly when determining the corrected date of birth for premature infants. It is unclear whether the information sheet capturing this data was given to the parent or guardian or whether the information was not submitted to the GAWM project. Explicit capture of prematurity has been addressed in ASQ Version 3.
- In one instance, an incorrectly calculated ASQ was later corrected through computerized calculations. Electronically calculated data fields minimize manual data calculation errors (for example, the participant's age in months, originally calculated manually was changed to an electronically calculated field, with a subsequent reduction in errors).
- In many instances, questionnaires submitted to the research study did not have the domain scores calculated or errors had been made in the transfer of data between the original questionnaire and summary sheet. Absence of immediate feedback and immediate feedback regarding ASQ findings was a concern expressed by parents in the telephone audits conducted in conjunction with this project. Calculation of summary scores at the time of the checkup provides a starting point for parental/guardian and health care provider discussion of expected growth and development, capacity building, and initiation of follow-up as required.
- Inconsistent wording of the questions between age-specific questionnaires resulted in the need for caution on the part of health care providers who transferred the information from the original questionnaire to the summary sheet. For example, in one age-specific questionnaire the

question refers to “vision concerns” while in another, the question is worded “vision ok”. This inconsistency has been addressed in ASQ Version 3.

- Reverse scoring of questions in the Meditech database in comparison to the original ASQ questions may have increased the risk of data entry error.
- Small sample sizes may have impacted the internal consistency findings.

### ***Opportunities for further research***

A further analysis of specific healthcare administrative agencies is strongly recommended in order to increase reliability across a larger scale and increase success of the program in the long term.

### ***Recommendations***

Based on the audit findings, it is recommended that:

- 1) The Grow Along with Me Steering Committee should:
  - a) Provide ongoing education regarding ASQ administration, scoring, data collection and follow-up process;
  - i) Highlight the correct response to developmental areas where the child previously performed, but no longer does because of the acquisition of more advanced skills (e.g. crawling replaced by walking). Squires et. al (1999) states “If parents mistakenly answer *not yet* or *sometimes* to an easier item... but *yes* to a more advanced item..., the score for the earlier item should be changed to 10 (for a *yes* response) before computing the total area score;” (p. 66)
  - b) Promote consistent use of developmental screening tools across all service providers;
  - c) Reinforce the calculation of ratio-based scores with staff responsible for scoring questionnaires;

- d) Stress the importance of Corrected Date of Birth (CDOB) when selecting age-specific questionnaires for premature infants;
  - e) Promote the collection and data entry of question level data in addition to summary data (including summary sheet, information sheet, and questionnaire). This would be particularly useful for future validity and reliability audits;
  - f) Encourage computerization of Ages and Stages data to reduce data transfer error and calculation error rates; and
  - g) Consider use of ASQ Version 3 in relation to identified issues that have been addressed in the newer version of the tool (e.g. consistency of wording across questionnaires, corrected date of birth).
- 2) ASQ developers should:
- a) Examine ASQ rigor and generalizability when used with a non-American, culturally diverse population.

## References

- Predictive Analytic Software (PASW). (2009). IBM Statistics Software Version 18. Retrieved Jan. 25, 2010 from the IBM Website :  
[http://www.spss.com/?source=homepage&hpzone=nav\\_bar](http://www.spss.com/?source=homepage&hpzone=nav_bar)
- Beaman, J., Huan, G., Beaman, T.C., Jeff, P. (2004, August 1). Tourism surveys: Sample size, accuracy, reliability, and acceptable error. *Journal of Travel Research*. 43(1), 67.
- Berg, K. & Latin, R. (2004) *Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation* (3rd edition). Lippincott Williams & Wilkins. Available through Culinary and Hospitality Industry Publications: Weimar, Texas 78962  
<http://www.chipsbooks.com/esscienc.htm>
- Garson, G.D. (2009). *Reliability Analysis*. Retrieved August, 12, 2009 from:  
<http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm>
- Grover, F., Hammermeister, K., and Shroyer, A. (1995). Quality initiatives and the power of the database. What they are and how they run. *Annals of Thoracic Surgery*. (60), 1614-21.
- Gwet, K. (2002). Computing inter-rater reliability with the SAS System. *Statistical Methods For Inter-Rater Reliability Assessment*, No. 3, October 2002.
- Morgan, G.A., Leech, N.L., Gloeckner, G.W., Barrett, K.C. (2007). *SPSS for Introductory Statistics. Use and interpretation (3rd Edition)*. Lawrence Erlbaum Associates: Mahwah, New Jersey.
- Squires, J., Potter, L. & Bricker, D. (1999). *The ASQ User's Guide for the Ages & Stages Questionnaires. A Parent-completed child monitoring system (Second edition)*. Paul. H. Brookes Publishing Co.: Baltimore, Maryland.

## Appendix A

### Code book for Validity Audit

#### Kappa alpha table in excel (A vs. B & B vs. C)

% Mismatch= percentage of inconsistent scoring

A= Total Score on the Questionnaire

B= Total Score on the Summary sheet

C= Total Score in the Master excel spreadsheet

Ref. # = Reference number of the questionnaire

Reason for mismatch = Reason for inconsistent scoring

# in the A, B, and C columns are the total score from all the domains (ex. 275=Communication 50, Gross motor 55, Fine motor 60, Problem solving 50, and Personal-social 60)

Yes = The information between A, B, and C were transferred correctly

No = The information between A, B, and C were not transferred correctly

Yes-incorrect = The information was transferred correctly from the questionnaire to the summary sheet to the master excel spreadsheet, but the information was calculated incorrectly

#### Individual Validity Table

Recorded Score= questionnaire recorded score (original source)

Q) missed= Question missed when calculating the score, or too many questions missed by parent while filling out the questionnaire

Incorrect Age= Miscalculation in the age of the child if premature

Miscalculation= incorrect calculating of the final score

#### Ratio Scores (See ASQ guidelines):

~Area total / number of items answered= ratio score

~Ratio score + previous total = final area score

#### Validated Domains for GAWM Table

Validated domains section has an original score and validated score. The next column will reflect whether the scores are the same (yes), inconsistent (no),

Miscalculation= Incorrect calculation within the summary sheet and the questionnaire

Sum miss= Miscalculated between the summary sheet and the score within the database. If a domain reflects the same scores but there is a Sum miss it is due to an inconsistency between the summary sheet and database score.

Missing info= Missing pages for the ASQ.

Miscalculation in database= Score was scored correctly on the questionnaire and the summary sheet yet an incorrect score was inputted into the database

Caught=Incorrect scoring was caught by the administrator who calculated the database scores and changed

Mark as 'yes'= A more complex developmental task has been answered as *yes* but a less complex developmental task the child may not do anymore is incorrectly marked as *sometimes* or *not yet*

## Guidelines for Scoring ASQs

Yes = 10

Sometimes = 5

Not Yet = 0

Correct an earlier question (Yes =10) if a more advanced question is answered yes or sometimes  
Scores that Fall **ON** or below the cut-off score are identified as needing a referral

Unanswered questions-Step 1-contact parents and correct or Step 2- Compute ratio score for 1-2  
questions unanswered

If 2 questions are unanswered double the ratio score before computing at end

Scored cannot be calculated if **more than 2** unanswered questions

If a number falls in between full digits .5 and higher –Round up, .49 and lower- round down

Assumption that the corrected date of birth is the same as the date of birth for questionnaires that  
have the information sheet filled out and the premature section left blank. If the information  
sheet was missing from the questionnaire the corrected date of birth was identified as  
inconclusive (999999) or (9999).

If the child was born **MORE THAN 3 WEEKS** prematurely the age must be adjusted to the  
corrected date of birth (CDOB) for up to 2 years

~CDOB=add the weeks of prematurity to the child's date of birth

Corrected date of birth (CDOB) = "A chronological date correction for weeks of prematurity  
when the actual date of birth is more than 3 weeks earlier than the expected birthdate. To  
calculate CDOB, add the weeks of prematurity to the child's date of birth" (Squires et al., 1999,  
p. 111).

Questionnaires are valid for 1 month before and 1 month after the indicated age (12 month can  
be used for 11-13 months)

If a questionnaire is not recorded as scored on each individual question, but the overall score for  
the domain is recorded there is an assumption that the unrecorded questions were answered and  
added up.

There is an assumption that if a score is scribbled out and a score is left clear and concise, that  
the parental guardian changed their answer to the clear answer.

If a question on the ASQ has two clear answers (not scribbled out) assume that the question was  
unanswered due to not being sure of the answer at the time

## Appendix B

### Definitions

**Internal Consistency:** 1. A common measure of reliability in the research literature is Cronbach's alpha. It is used to assess the internal consistency reliability of several items or score that the researcher wants to add together to get a summary or summated score. Alpha is based on a correlation matrix. Alphas should be positive and usually greater than .70 to provide good support for internal consistency reliability. (Morgan, Leech, Gloeckner & Barrett, 2007, p. 129)

2. All of the data for a given observation appear to be feasible if examined together (e.g. ASQ questions relate to the developmental scores and, in turn the domain scores relate to the overall ASQ score).

**Reliability:** According to Squires et al. (1999) reliability is the consistency of the test scores and the extent to which a generalization can be made from a test result by different observers and differing times (p. 112). According to Berg & Latin (2004) "reliability measures the consistency or repeatability of test scores or data." (p.165) Kappa Alpha (Crosstabs) Reliability is the most common used measurement tool for internal consistency (Garson, 2009). Kappa-Alpha reliability is used to examine the measure of agreement scorer reliability. When there are two nominal variables with the same values (usually scores using the same codes or two different raters), Cohen's kappa alpha is computed in order to check the agreement between measures or reliability (Gwet, 2002). The reliability within the audit represents the concordance of the ASQs individual scores, scoring summary sheets and ASQ database.

**Validity:** The degree to which a research variable reflects what it is intended to measure (e.g. the extent to which a condition mentioned in the database reflects the condition of the variable of interest [e.g. the growth and development status of the child]).