



Options for Measuring Changes Over Time on the ASQ-3

June 2017



LeCroy & Milligan
ASSOCIATES, INC.

Options for Measuring Changes Over Time on the ASQ-3 June 2017

Submitted to:

Chapin Hall
University of Chicago
1313 E 60th Street
Chicago, IL 60637
Ph: (773) 256-5127
www.chapinhall.org

Submitted by:

LeCroy & Milligan Associates, Inc.
2002 N. Forbes Blvd. Suite 108
Tucson, AZ 85745
Ph: (520) 326-5154
Fax: (520) 326-5155
www.lecroymilligan.com



Acknowledgments:

The research team for this project would like to thank Chapin Hall for the opportunity to investigate options for the ASQ-3. The research team includes Darlene Lopez, MA, ABD, Craig LeCroy, PhD., Michele Schmidt, MPA, and Greg Goodman, MS.

About LeCroy & Milligan Associates:

Founded in 1991, LeCroy & Milligan Associates, Inc. is a consulting firm specializing in social services and education program evaluation and training that is comprehensive, research-driven and useful. Our goal is to provide effective program evaluation and training that enables stakeholders to document outcomes, provide accountability, and engage in continuous program improvement. With central offices located in Tucson, Arizona, LeCroy & Milligan Associates has worked at the local, state and national level with a broad spectrum of social services, criminal justice, education and behavioral health programs.

Suggested Citation:

Options for Measuring Changes Over Time on the ASQ-3, LeCroy & Milligan Associates, Inc. (2017).



Introduction

The purpose of this study is to determine the potential of the Ages and Stages Questionnaire, Third Edition (ASQ-3, <http://www.agesandstages.com>) as an outcome measure for child development. The ASQ-3 is designed as a screening tool for developmental delays to be used with children aged 2 months to 60 months (Squires, et al., 2009). The ASQ-3 is used by many home visiting programs (<http://agesandstages.com/about-asq/who-uses-asq/>) as it is an easy tool to use with parents and gives parents an opportunity to be directly involved in helping determine the developmental level of their child. Other child development tools often require trained observers and are costlier in terms of licensing and personnel time. This study in conjunction with other research can be used to determine the potential use of the ASQ-3 for tracking a child's developmental progress in addition to its use as a screening tool.

Methods

The ASQ-3 questionnaire consists of 30 developmental items in five domains of child development: communication, gross motor, fine motor, problem solving and personal-social. There are six questions in each domain and the parents indicate "yes" (10 points), "sometimes" (5 points) or "not yet" (0 points) for their child's ability to perform a task. Each domain score is the sum of the individual items with a mean adjusted value for one or two missing questions within a domain. The domain score is then compared with established screening cut-off points, and is coded as "typical", meaning there are no developmental concerns, "questionable", meaning that the child should be monitored for possible delays, and "delayed", meaning that a refer should be given for the child for developmental assistance. In addition to the five domains, there is an additional set of questions called "Overall" with yes/no questions and a place for parents and providers to write additional comments. The overall section provides useful information when referrals are made, but is not included in the scoring of the domains or ASQ-3 as a whole.

This study used ASQ-3 scores from 215 children given at 6, 9, and 12 months of age who were receiving home visiting services through Healthy Families Arizona. As part of the agreement to use the data for this study, no demographic information was used in this study, only an identification number to match the 6, 9, and 12 month ASQ-3 forms to a specific record. The ASQ-3 forms received are exactly as



the parent completed them, no trained staff or researchers collected the data. Because the forms are completed by the parents and not a professional, there were forms where some parents skipped questions. If only one or two questions in a domain were missing the mean of the other questions was used to replace the skipped questions to create a domain score as described by the ASQ-3 User Manual (Squires, et al., 2009). For domains with three or more skipped questions, the domain score was not calculated. See Table 1 for a description of the missing data, and useable sample sizes for each domain and time point of the ASQ-3.

Table 1: Raw Domain Score Means, Standard Deviations, and Sample sizes by time point

Domain	6 Month ASQ-3 Mean (St. Dev) N	9 Month ASQ-3 Mean (St. Dev) N	12 Month ASQ-3 Mean (St. Dev) N
Communication	52.57 (7.31) N= 214	48.24 (11.25) N= 207	50.61 (10.20) N= 215
Gross Motor	47.94 (10.37) N= 214	44.47 (14.31) N= 214	52.12 (11.18) N= 213
Fine Motor	52.28 (9.65) N= 211	54.14 (8.28) N= 215	54.80 (7.26) N= 213
Problem Solving	54.62 (8.41) N= 210	51.68 (9.20) N= 208	51.31 (8.82) N= 214
Personal-Social	52.51 (9.34) N= 213	46.27 (10.25) N= 208	49.00 (10.19) N= 214

As we had access to the full ASQ-3 scores, the data included the individual question responses, the domain scores, and the overall questions and comments. The domain outcomes of typical, questionable, or delayed were calculated using the published cut-off scores for each domain as well as for the screen as a whole. An ASQ-3 screen is considered delayed if at least one domain scores below the cut-off, and is considered questionable if at least one domain falls within the questionable or monitoring zone.

The cut-off scores for questionable or delayed screens are based off of the standard deviations of the normed sample used during the development of the ASQ-3 as described in the ASQ-3 User Manual. This makes it difficult to directly compare raw domain scores between versions of the ASQ-3 as a score of 30 at one time point might be considered typical and at the next time point considered to be in the



questionable or delayed zone. Using the means and standard deviations for each ASQ-3 time period, standardized scores were created for each domain. These standardized scores were then used to look at differences in scores between ASQ-3 versions.

There are also six questions which remain the same between the 6 month and 9 month versions, and sixteen between the 9 month and 12-month version of the ASQ-3. The research team thought that it might be potentially useful to analyze the changes over time for individual questions. Hypothesizing that there should be a relationship between changes in individual question scores and domain outcomes, difference scores were calculated for these matching questions and the correlations between the difference and domain outcomes were compared.

The additional section of the ASQ-3 known as “Overall” has also not previously been reported on in published studies. The research team was interested in learning if the responses to the “Overall” section were related to the outcome of typical, questionable, or delayed as well as to specific domains. The responses to the “Overall” section of the ASQ-3 were compared between children with and without concerns using t-tests, and the relationship with the standardized scores within domains was reviewed with correlational analyses.

Two research questions guided this study:

- Primary research question: Can the ASQ-3 as it is used by home visitation programs be used to measure changes over time? Using the following:
 - Screening outcomes of typical, questionable, or delayed;
 - Standardized scores for each ASQ-3 domain; and
 - Matching questions between versions.
- Secondary research question: Does the “Overall” section of the ASQ-3, a section not used for the determination of screening outcomes, provide additional useful information related to outcomes.

All analyses were conducted using SPSS version 22.



Results

ASQ-3 Cut-off Outcomes

A child screens as typical, questionable, or delayed in each domain based on the cut-off scores provided in the ASQ-3 User Manual. If a child scores less than the referral cut-off score in one or more domains the ASQ-3 outcome is denoted as delayed and a referral for services is recommended (See Table 4.4 in the ASQ-3 User Manual for reference). If a child has no domains that score below the cut-off, but at least one is in the monitoring zone they ASQ-3 outcome is classified at questionable and additional monitoring is recommended. Using the ASQ-3 screening outcomes of typical, questionable, and delayed we analyzed the changes over time from 6 months to 9 and 12 months using chi-square analyses to determine distribution patterns and Spearman's Rho (non-parametric) correlations to observe potential relationships.

Eleven children had a delay in at least one ASQ-3 domain at 6 months. Of those 11 children:

- 3 had delays at both 9 and 12 months as well;
- 3 had delays at 9 months, but not at 12 months;
- 1 was questionable at both 9 and 12 months.
- 2 were questionable at 9 months, but were typical at 12 months; and
- 2 had delays at 12 months, but were questionable at 9 months.

Thirty additional children screened in as questionable on at least one domain at 6 months. Of those 30 children:

- 8 were found typical at 9 and 12 months;
- 7 remained questionable at both 9 and 12 months
- 7 remained questionable at 9 months, but were typical at 12 months;
- 2 showed delays at 9 months, but were typical at 12 months;
- 3 showed delays at 9 months, but were questionable at 12 months;
- 2 showed delays at 12 months, but were questionable at 9 months;
- 1 showed delays at both the 9 and 12 months times.



An additional 7 children screened as delayed at 9 months. Of those 7 children:

- 3 continued to show delays at 12 months;
- 4 screened as typical at 12 months.

A total of 28 children screened as delayed at one of the 3 time points, 13% of the sample. Chi-square analyses were unable to show significant differences in distribution. The small number of children with delays make it difficult to interpret the meaningfulness of this observation, or to analyze differences within the specific domains of Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social. However, correlational analyses do show that there is a relationship between the screening outcomes: Spearman's Rho = 0.466 between 6 and 9 months; 0.354 between 6 and 12 months; and 0.406 between 9 and 12 months all at $p=0.000$. Although this study was unable to identify trends in the distribution of screening outcomes, the statistically significant relationship between the screening outcomes at each time point suggests that it may be possible to use the screening outcomes for outcomes analyses in data where a sufficient number of children screen as delayed or questionable.

Standardized ASQ-3 Scores

As discussed in the methods section above, the cut-off scores for questionable and delayed screening outcomes are different for each time point of the ASQ-3. Due to this difference, it is not appropriate to directly compare the raw scores across ASQ-3 administrations. Standardized scores were created for each domain so that comparisons could be made between the different ASQ-3 versions. In order to determine trends over time from the 6 month to the 9 month to the 12-month administration of the ASQ-3, repeated measures analyses were conducted.

The repeated measures analyses were found to be significant for all domains except for Personal-Social. This indicates that there are statistically significant differences between the time points, however, the differences are not necessarily linear. The ASQ-3 Communication domain has a peak at 9 months, Gross Motor and Fine Motor have valleys at 9 months compared to the 6 and 12 month scores, while the Problem Solving and Personal-Social show a downward slope from 6 to 9 to 12 months. This indicates that developmental changes over time follow different pattern for each domain. The standardized scores provide sufficient information to be able to determine changes over time through repeated measures or pairwise comparisons. However, based on the differential patterns for each domain, it is recommended that domain specific analyses are conducted. The means and standard deviations for each domain along with the F-statistic are listed in Table 2.



Table 2: Standardized ASQ-3 Scores Means and Standard Deviations by Month

Domain	Mean	Standard Deviation	Outcome
Communication			
6 month	0.369	0.744	F (2, 204)=19.911, p=0.000
9 month	0.784	0.915	
12 month	0.534	0.743	
Gross Motor			
6 month	0.195	0.881	F (2, 209)=24.316, p=0.000
9 month	-0.167	0.992	
12 month	0.156	0.790	
Fine Motor			
6 month	0.291	0.792	F (2, 207)=3.114, p=0.047
9 month	0.168	0.794	
12 month	0.286	0.823	
Problem Solving			
6 month	0.374	0.739	F (2, 201)=5.477, p=0.005
9 month	0.213	0.843	
12 month	0.189	0.817	
Personal-Social			
6 month	0.371	0.814	F (2, 204)=0.765, p=0.467
9 month	0.332	0.868	
12 month	0.287	0.834	



Matching Questions

Having access to the individual question scores allowed for an additional analysis to be conducted to determine if questions that remained the same between versions of the ASQ-3 had a relationship to the screening outcomes and could possibly be used to provide additional information about a child's developmental growth over time. Six questions remain the same from 6-month version of the ASQ-3 to the 9-month version, as well 16 different questions that remain the same from the 9-month to the 12-month version. No questions are the same during all three time periods.

Three different analyses were conducted to determine if the raw difference between scores on questions asked at two time points could be used to show changes. First paired sample t-tests were used to determine significant differences between the two time points for each question. All of the question pairs were found to have statistically significant differences. The results are shown in Table 3.

The second analysis used independent samples t-tests to compare differences between children who screened as typical versus those that screened as questionable or delayed. Ten of the questions were significantly different between the two groups. The results of the statistically significant differences are shown in Table 4.

The final analysis looked at the relationship between the difference scores on the matching questions and if the ASQ-3 was determined to be typical, questionable, or delayed for the corresponding domain at the second time period. Spearman Rho correlation coefficients were found to be statistically significant for four question pairs:

- Communication at 9 and 12 months for "Does your baby say three words, such as "Mama," "Dada," and "Baba"?" (Spearman's Rho = 0.173, $p=0.013$)
- Gross Motor at 9 and 12 months for "Does your baby walk beside furniture while holding on with only one hand?" (Spearman's Rho = 0.171, $p=0.012$)
- Fine Motor at 6 and 9 months for "Does your baby pick up a small toy with only one hand?" (Spearman's Rho = -0.165, $p=0.017$)
- Problem Solving at 9 and 12 months for "While holding a small toy in each hand, does your baby clap the toys together?" (Spearman's Rho = -0.188, $p=0.007$)



Table 3: Comparisons Between ASQ-3 Matching Questions

	Mean	Standard Deviation	t	df	p
Does your baby make sounds like “da,” “ga,” “ka,” and “ba”? 6 to 9 month	-2.1117	4.2233	-7.176	205	0.000
If you copy the sounds your baby makes, does your baby repeat the same sounds back to you? 6 to 9 month	-2.0146	4.0958	-7.060	205	0.000
Does your baby make two similar sounds like “ba-ba,” “da-da,” or “ga-ga”? 9 to 12 month	-0.7282	2.4101	-4.336	205	0.000
If you ask your baby to, does he play at least one nursery game even if you don’t show him the activity yourself? 9 to 12 month	-1.4078	3.9458	-5.121	205	0.000
Does your baby follow one simple command, such as “Come here,” “Give it to me,” or “Put it back,” without your using gestures? 9 to 12 month	-1.8841	3.9198	-6.915	206	0.000
Does your baby say three words, such as “Mama,” “Dada,” and “Baba”? 9 to 12 month	-2.4146	4.4169	-7.827	204	0.000
If you hold both hand just to balance your baby, does she support her own weight while standing? 6 to 9 month	-0.6103	2.5887	-3.441	212	0.001
While holding onto furniture, does your baby bend down and pick up a toy from the floor and then return to a standing position? 9 to 12 month	-3.5514	4.5455	-11.429	213	0.000
While holding onto furniture, does your baby lower himself with control (without falling or flopping down)? 9 to 12 month	-3.4742	4.3595	-11.631	212	0.000
Does your baby walk beside furniture while holding on with only one hand? 9 to 12 month	-4.8592	4.6306	-15.315	212	0.000
Does your baby pick up a small toy with only one hand? 6 to 9 month	-0.6161	2.0353	-4.397	210	0.000
After one or two tries, does your baby pick up a piece of string with her first finger and thumb? 9 to 12 month	-0.6338	2.8229	-3.277	212	0.001
Does your baby pick up a crumb or Cheerio with the tips of his thumb and a finger? 9 to 12 month	-0.9155	2.6569	-5.029	212	0.000
Does your baby put a small toy down, without dropping it, and then take her hand off the toy? 9 to 12 month	-1.6667	4.1113	-5.916	212	0.000
Does your baby pass a toy back and forth from one hand to the other? 6 to 9 month	-1.4706	3.6191	-5.804	203	0.000
While holding a small toy in each hand, does your baby clap the toys together? 9 to 12 month	-1.4493	3.7295	-5.591	206	0.000
Does your baby poke at or try to get a crumb or Cheerio that is inside a clear bottle? 9 to 12 month	-2.0773	4.4649	-6.694	206	0.000
After watching you hide a small toy under a piece of paper or cloth, does your baby find it? 9 to 12 month	-1.4010	3.6826	-5.473	206	0.000
While your baby is on her back, does she put her foot in her mouth? 6 to 9 month	-1.0386	3.8489	-3.883	206	0.000
When you hold out your hand and ask for her toy, does your baby offer it to you even if she doesn’t let go of it? 9 to 12 month	-2.5962	4.3395	-8.628	207	0.000
When you dress your baby, does he push his arm through a sleeve once his arm is started in the hole of the sleeve? 9 to 12 month	-1.2379	4.5582	-3.898	205	0.000
When you hold out your hand and ask for her toy, does your baby let go of it into your hand? 9 to 12 month	-3.8164	5.0411	-10.892	206	0.000



Table 4: Comparisons Between ASQ-3 Matching Questions by Screening Outcome

Similar question differences by concerns		N	Mean	Std. Deviation	t(df), p (2-tailed)
Does your baby make sounds like “da,” “ga,” “ka,” and “ba”? Difference from 6 to 9 months	Typical	108	1.574	3.529	t (176.3) = -1.902, p = 0.059
	Questionable or Delayed	98	2.704	4.825	
Does your baby follow one simple command, such as “Come here,” “Give it to me,” or “Put it back,” without your using gestures? Difference from 9 to 12 months	Typical	109	1.009	3.102	t (170.1) = -3.412, p = 0.001
	Questionable or Delayed	98	2.857	4.484	
While holding onto furniture, does your baby bend down and pick up a toy from the floor and then return to a standing position? Difference from 9 to 12 months	Typical	113	1.770	3.655	t (189.8) = -6.565, p = 0.000
	Questionable or Delayed	101	5.545	4.631	
While holding onto furniture, does your baby lower himself with control (without falling or flopping down)? Difference from 9 to 12 months	Typical	113	1.947	3.370	t (176.9) = -5.730 p = 0.000
	Questionable or Delayed	100	5.200	4.710	
Does your baby walk beside furniture while holding on with only one hand? Difference from 9 to 12 months	Typical	112	3.973	4.657	t (211.0) = -2.995 p = 0.003
	Questionable or Delayed	101	5.842	4.420	
Does your baby pick up a small toy with only one hand? Difference from 6 to 9 months	Typical	111	0.270	1.321	t (144.6) = -2.558 p = 0.012
	Questionable or Delayed	100	1.000	2.562	
After one or two tries, does your baby pick up a piece of string with her first finger and thumb? Difference from 9 to 12 months	Typical	112	0.223	1.558	t (131.3) = -2.176 p = 0.031
	Questionable or Delayed	101	1.089	3.715	
Does your baby pick up a crumb or Cheerio with the tips of his thumb and a finger? Difference from 9 to 12 months	Typical	112	0.402	1.521	t (134.8) = -2.925 p = 0.004
	Questionable or Delayed	101	1.485	3.431	
After watching you hide a small toy under a piece of paper or cloth, does your baby find it? Difference from 9 to 12 months	Typical	109	0.826	2.928	t (168.5) = -2.351 p = 0.020
	Questionable or Delayed	98	2.041	4.298	
While your baby is on her back, does she put her foot in her mouth? Difference from 6 to 9 months	Typical	108	0.463	3.521	t (194.0) = -2.255 p = 0.025
	Questionable or Delayed	99	1.667	4.103	
When you hold out your hand and ask for her toy, does your baby offer it to you even if she doesn’t let go of it? Difference from 9 to 12 months	Typical	109	1.743	3.560	t (177.3) = -2.988 p = 0.003
	Questionable or Delayed	99	3.535	4.910	

The paired questions provide a way to differentiate between changes over time on specific developmental questions and may provide insight into specific areas of child development outcomes. However, only 4 of the 22 pairs analyzed shows statistically significant relationships with domain outcomes of the later time period.



Overall Questions

The usefulness of the “Overall” questions to determine outcomes was the secondary question in this study. In order to determine the potential usefulness of the “Overall” questions, the distribution of Yes/No responses within the ASQ-3 screening outcome of typical, questionable, or delayed was analyzed using chi-squared analyses. At each of the three ASQ-3 time points very few parents responded with answers that might indicate an issue or concern. However, two questions showed statistically significant differences in distributions:

- When you help your baby stand, are his feet flat on the surface most of the time? At 6 months: $\chi = 10.967, p = 0.004$. At 9 months: $\chi = 9.197, p = 0.010$. At 12 months: $\chi = 16.933, p = 0.000$.
- Does anything about your baby worry you? At 9 months $\chi = 6.149, p = 0.046$. At 12 months: $\chi = 14.871, p = 0.001$

Children that screened as delayed had more parents indicate that the child’s feet were not flat when standing, and expressed more concerns in the “Overall” questions. However, the written comments that accompany these questions were not consistently completed by parents. For the parents who did respond the answer to “feet flat on the floor” was most commonly indicating that the baby was standing on their toes. For the question on worry, the answers ranged from concerns about behavior and fussiness, to recent illness, to issues with walking or crawling. These two questions appear to have a strong relationship to the screening outcome; however, the quality of the comments is inconsistent and may not provide sufficient information to guide decision making or determine outcomes.

Limitations

There are several limitations to this study. The small number of children screened as delayed made analyses difficult to find statistically significant. A larger sample of children that includes a greater number of children screened as delayed is necessary for sufficient power to be able to determine differences using the screening outcomes of typical, questionable, or delayed. The comment sections in the “Overall” questions were also not consistently completed by parents, or provided little useful additional information.

This study also does not include a control group. All children within this study had been participating in Healthy Families Arizona services so there is no counterfactual.



Summary

Using the ASQ-3 screening outcomes of typical, questionable, and delayed alone may be limited in showing developmental changes over time. The reason for this is two-fold:

- There are few children that fall within the delayed and monitoring zones and there is a lot of fluctuation over time at least at the 6, 9, and 12 month ASQ-3. More distal comparisons may be more useful, and might provide additional options for future studies.
- The standardized scores appear to be a more useful measure of change over time within the specific developmental domains. An advantage of the standardized scores is that they give a more accurate indication of the extent of delays when delays are present. The standardized scores may be useful for tracking developmental progress over multiple time points.

The matched questions between ASQ-3 versions do show differences over time so may be useful for looking at specific questions within a developmental domain. For example, if a home visitation program has been working with a family on a specific fine motor skill such as having the child pick up a string, the matching question on the 9 and 12 month ASQ-3 could be used to determine the success of that intervention. However, caution is advised in using the matched questions for overall changes as only 4 of the 22 paired questions analyzed showed significant relationships with ASQ-3 screening outcomes.

Finally, for the “Overall” questions, two questions appear to be related to the screening outcomes. However, the qualitative comments are inconsistent and may not provide additional useful information. Further research into the usefulness of “Overall” question section and parental comments is recommended.

Based on the findings of this study, the ASQ-3 appears to be useful for tracking a child’s developmental changes over time. The results of the study suggest that obtaining the ASQ-3 domain raw scores and converting them to standardized scores provides the best option for analysis. However, the statistically significant correlational relationships for the screening outcomes indicates that they may also be used to make more global statements about change over time.



References

Squires, J., Bricker, D., Twombly, E., & Potter, L. (2009). *ASQ-3 User's Guide*. Baltimore: Paul H. Brookes Publishing Co.

