

Report on a Study of the Reliability and Validity of the Quality Experience Scanning Tools (QUEST)

**Prepared for: The High Five Program
Parks and Recreation Ontario**

June 12, 1998

LCP
THE LEVY-COUGHIN PARTNERSHIP
Market & Social Research

Table of Contents

Summary of Findings	i
1.0 Introduction to Quality Assurance in Children's Recreation	2
1.1 Current Project Status	2
1.1.1 Phase I	2
1.1.2 Phase II	3
1.1.3 Phase III	3
1.2 Purpose of this Report	4
2.0 Introduction to QUEST	5
2.1 Observing the Child's Experience - Comprehensive (QUEST2)	5
2.2 Overview of the Approach Used for the Reliability/Validity Study of QUEST2	6
2.2.1 Internal Consistency and Inter-Rater Reliability of QUEST2	6
2.2.2 Orientation	6
2.2.3 First Pilot Test	7
2.2.4 Second Pilot Test	8
2.2.5 Design of the Reliability Assessment	8
2.3 Construct Validity	10
2.3.1 Content Validity	10
2.4 The Review of Best Practices (QUEST1)	11
2.4.1 Overview of the Validity Assessment Process of QUEST1	12
3.0 Analysis of the Scales: Observing the Child's Experience (QUEST2)	13
3.1 Description of QUEST2 Scales	13
3.2 Results of the Initial Reliability and Validity Study of QUEST2	14
3.3 Internal Consistency Reliability	14
3.3.1 The Leader/Child Interactions Scale	14
3.3.2 The Supervision/Safety Scale	16
3.3.3 The Child/Peer Interactions Scale	17
3.3.4 The Leader Behaviour and Interactions Scale	18
3.3.5 The Program/Activities Scale	19
3.3.6 The Environment/Facility Scale	20
3.3.7 The Equipment/Materials Scale	21
3.3.8 The Total Score	22
3.4 Inter-Rater Reliability of QUEST2	23
3.5 Construct Validity of QUEST2	24
3.5.1 Leader/Child Ratios	24
3.5.2 Years of Education of Leaders	24
3.6 Content Validity of QUEST2	26
3.7 Summary of Initial Findings QUEST2	28
3.8 Content Validity of QUEST1	29
3.9 Preliminary Conclusions/Discussion Regarding QUEST2	30
3.9.1 Discussion	30
3.9.2 Specific Items	30
3.9.3 Interrater Reliability	31

	3.9.4	The Length of the Tool	31
	3.10	Conclusions	32
4.0		Additional Analyses and QUEST2 Modifications	33
	4.1	The Redefined Safety/Supervision Scale	35
	4.2	The Program Characteristics/Supports Scale	35
	4.3	The Redefined Total Score	36
	4.4	Content and Construct Validity of the Redefined Scales	36
	4.5	Interrater Reliability of the Redefined Scales	37
	4.6	Summary of Findings of the Redefined Version of QUEST2	39
	4.7	The Shorter Version of QUEST2	39
	4.7.1	Administration Time	40
	4.7.2	The Items Selected for QUEST3	40
	4.7.3	Psychometric Properties of QUEST3	41
5.0		Scoring QUEST2	42
	5.1	Sub-Totals for Each Scale/Category and Overall Total for QUEST2	42
	5.2	Item by Item Scoring	44
	5.3	Target Scores for High Five "Accreditation"	45
	5.4	Discussion on the Administration of QUEST3 and QUEST2	45
	5.5	Norms for QUEST2 and QUEST3	46
	5.6	Final Statement	46
Appendix A		Observing the Child's Experience - Comprehensive (Under Separate Cover)	47
Appendix B		Review of Best Practices (Under Separate Cover)	48
Appendix C		Average Ratings of QUEST1 Items by Recreation Providers	49

Summary of Findings

This document presents the results of study investigating the reliability and validity of the Quality Experience Scanning Tools (QUEST) of the High Five Program.

The QUEST package is composed of three instruments:

- QUEST1: Review of Best Practices;
- QUEST2: Observing the Child's Experience - Comprehensive; and
- QUEST3: Quick Scan (a shorter version of QUEST2).

The reliability of the QUEST2 and QUEST3 was assessed using two methods.

Internal consistency reliability was determined using Cronbach's Alpha. Interrater reliability was assessed using the intraclass correlation and the Spearman Brown Prophecy formula adjusted for the number of raters.

The construct validity of QUEST2 and QUEST3 was assessed by correlating their scale and total scores with leader/child ratio and years of education of leaders. The correlation of the total score of QUEST3 with the total score of QUEST2 was also examined.

The content validity of QUEST1 and QUEST2 was determined by requesting a sample of experts to rate the importance of each item as a factor effecting the healthy development of children.

Method

Reliability and Construct Validity

A female and male rater were trained in the use of QUEST2. The raters independently observed and rated 36 programs representing a cross-section of the program types to which the tools were to be applied. They also collected information on leader/child ratios and leader education.

Content Validity

A sample of experts were sent a questionnaire requesting them to rate the importance of each item in QUEST2 as a factor effecting the healthy development of children. The experts used a scale ranging from one to five, where one was not at all important and five was extremely important.

A second sample of experienced recreation practitioners were sent a questionnaire asking them to rate the items in QUEST1 in a similar way.

Results

QUEST1

Content Validity

With respect to the **individual items** in QUEST1, eighty eight percent were rated four or higher by practitioners, and 53% were rated 4.5 or higher. Eighty nine percent of the **sections** of QUEST1 (16 out of 18) had an average rating of four or higher. In fact, half of the sections were rated 4.5 or higher. Overall, the providers rated all sections at an average of 4.33. According to the recreation providers, QUEST1 has excellent content validity.

QUEST2

After interpretation of the initial findings, modifications were made to QUEST2. The table below presents the results of the reliability and validity assessment of QUEST2.

Summary of Reliability/Validity of the QUEST2

Scale	Internal Consistency Reliability		Spearman Brown Prophecy		Content Validity ¹	Construct Validity	
	Alpha	Intraclass Correlation (ICC 3,1) One Rater	Two Raters	Three Raters		Expert Ratings	Staff/Child Ratio
Leader/Child Interactions	0.96	.78	.88	.92	4.57	-0.26	0.29
Supervision/Safety	0.73	.80	.89	.92	4.47	-0.46	0.46
Child/Peer Interactions	0.76	.65	.79	.85	4.47	-0.23	0.23
Leader Behaviour and Interactions	0.82	.95	.97	.98	4.44	-0.40	0.66
Program Characteristics/ Supports	0.85	.62	.76	.83	4.25	-0.24	0.33
Total Score	0.90	.91	.95	.97	4.44	-0.41	0.62

Reliability

Internal Consistency Reliability

The alphas for the QUEST2 scales ranged from .73 to .96 indicating good to excellent internal consistency reliability.

¹

Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

Interrater Reliability

The intraclass correlations ranged from .62 to .95 assuming one rater scores a program; from .76 to .97 assuming two raters score a program and average their ratings; and from .83 to .98 assuming three raters score the program.

Given the range and complexity of programs in which QUEST2 will be applied, we strongly recommend that during the initial stages of the initiative ***the tool be administered by at least two raters and that the final scores for the tool be calculated as the average of their ratings.*** If the tool is administered in this way, the interrater reliability indices will range from good to excellent.

Content Validity

Twenty six of the 27 items in QUEST2 (96%) had an average rating of 4 or higher. The average overall rating for all items was 4.4. According to the opinions of the experts in child development, QUEST2 has excellent content validity.

Construct Validity

The scales and total scores have moderate and in some cases significant⁵ negative correlations with staff to child ratios. We also found moderate, and in some cases significant correlations between years of education and the scores of QUEST2.

Given that both measures of the construct validity of QUEST2 revealed relationships that were in the "right" direction, and considering the range of programs and the size of the sample, these results indicate that QUEST2 has a degree of construct validity to support its use.

Summary

The following scales of QUEST2 appear to have sufficient internal consistency reliability, interrater reliability, construct and content validity to support their use as measures of factors effecting the healthy development and well-being of children in sport/recreation programs:

leader/child interactions;
supervision/safety;
leader behaviour and interactions;
child/peer interactions; and
program characteristics and supports.

5

At the 95% level of confidence.

QUEST3

A shorter version of QUEST2 was developed called QUEST3 which contained 14 items. The table below presents the results for QUEST3 as a single scale.

Summary of Reliability/Validity of QUEST3

Scale	Internal Consistency Reliability				Content Validity ⁶	Construct Validity		
	Alpha	Intraclass Correlation (ICC 3,1) One Rater	Spearman Brown Prophecy Two Raters	Spearman Brown Prophecy Three Raters	Expert Ratings	Staff/Child Ratio	Education	Correlation with QUEST2 Total Score
QUEST3	0.93	0.88	0.93	0.96	4.60	-0.39	0.46	0.98

Findings

The scale produces an alpha of .93 and its intraclass correlations ranges from .88 for one rater to .96 for three. The average ratings by experts is 4.6, and its construct validity is acceptable.

In addition, the total score for QUEST3 was correlated with the total score for QUEST2. The correlation between the two scores is .98, indicating that the QUEST3 total score is an excellent predictor of the QUEST2 total score.

Summary

QUEST3 has sufficient internal consistency reliability, interrater reliability, construct and content validity to support its use as a single scale measure of factors effecting the healthy development and well-being of children in sport/recreation programs.

We would recommend that the internal consistency reliability, interrater reliability, and construct validity of QUEST continue to be investigated and reviewed on a regular basis as the program unfolds.

6

Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

1.0 Introduction to Quality Assurance in Children's Recreation

Quality Assurance in Children's Recreation and Sport is a process still in its developmental stages. The intent of the process is to ultimately impact on the healthy development of children in Ontario by ensuring a high standard of quality in children's' recreation and sport programs in all venues throughout the province. As a non-regulated field, recreation and sport has traditionally governed itself, with each service provider establishing its own processes for quality assurance.

Parks and Recreation Ontario (PRO), as the central body representing recreation, believes that despite the efforts and success of the various service providers, there is a need for an overriding and consistent assurance of quality within the field as a whole. To this end, Quality Assurance in Children's Recreation and Sport is a process committed to assisting children along the path of healthy development by:

developing and maintaining a high level of knowledge and expertise in child development among practitioners;

assisting parents with making informed choices; and

providing practitioners with the tools for enhancing and maintaining a high level of quality.

1.1 Current Project Status

1.1.1 Phase I

Phase I of the project was completed in December 1995. During this phase several products were created to drive the quality assurance process forward. These included:

A: Our Commitment to Children for Healthy Development In Recreation Settings

B: Desired Outcomes for Children in Recreation and Sport

These principles and objectives were created to serve as a guideline for program development and assessment.

The Model for Quality Assurance

The central goal of this model is quality programs for healthy child development. It is a voluntary compliance model, with public expectations driving standards and quality. Four components influence each other in the achievement of this goal.

1 Public Awareness-Education

Recreation and sport are closely connected with the communities and families they serve and consequently are driven by community needs. Parents require education to understand how

recreation and sport should be enhancing children's development in order to be able to take responsibility when this is not occurring. Public education and high expectations, coupled with a market driven environment will prompt service providers to make appropriate changes.

2 Staff Training

For quality assurance to occur, there needs to be a consistent standard of knowledge focussing on child development and application in recreation and sport. The Quality Assurance Model includes the development and delivery of a central training program for all practitioners working with children.

3 Communication Plan

It is also important that all pertinent audiences receive communication and are mobilized appropriately. A detailed Communication Plan has been developed to serve as a guideline in this process.

4 Quality Assessment Measurement

The measurement process is designed to provide practitioners with the ability and knowledge to:

- review standards and assess programs;
- educate parents and staff; and
- plan high quality programs

The process is currently in prototype form and includes two tools -the Review of Best Practices (QUEST1) and the Quality Experience Scanning Tool (QUEST2).

QUEST1 and QUEST2 were developed after a thorough review of the literature related to factors effecting the healthy development of children, as well as extensive consultation with experts in child development.⁵

1.1.2 Phase II

Phase II entails:

1 Vision and Direction Paper

A Vision and Direction Paper is being created to guide the quality assurance process into the future.

⁵

Please see, "Final Report, Phase I, The Quality Assurance Standards Project. Parks and Recreation Ontario, 1996" for details related to the tools' development.

2 Implementation of the Quality Assurance Model

Various products, resources and aids are being created to assist parents with their choices, as well as assisting practitioners. The development of three training modules for staff and the delivery of training will be completed.

3 Development and implementation of a Marketing Strategy

Using the communication Plan, a marketing strategy is being developed to address all pertinent audiences. One key outcome of this plan is the naming of the overall initiative to be the **High Five Program**.

4 Design of Management and Delivery Systems

Under the PRO umbrella, a "Children's Council" Management structure as a delivery system for all products is being developed.

5 Development of Partnerships-Internal and External

A very important element of Phase II is the development of a strategy to mobilize the sector to ensure "buy-in" to the process, create partnerships with other stakeholders and secure financial support from the private sector.

1.1.3 Phase III

Phase III entails:

- Development of the training;
- Delivery of the training;
- Marketing;
- Development of the Guide to Best Practices;
- Development of the Sport Specific tool;
- Production of newsletters;
- Production of the Parent handbook;
- Production of the Staff handbook; and
- In-depth validation of QUEST.

1.2 Purpose of this Report

In relation to the last task above, this document presents the results of a study investigating the reliability and validity of QUEST. The next chapter details the methodology employed for the study. Chapter three presents the results as well as a summary and preliminary conclusions. Chapter four details the changes made to QUEST2 based on the findings, and chapter five addresses issues related to the administration and scoring of the tools.

2.0 Introduction to QUEST

The complete measurement package developed for the High Five Program is called the Quality Experience Scanning Tools (QUEST). The package is composed of three instruments:

- QUEST1: Review of Best Practices;
- QUEST2: Observing the Child's Experience - Comprehensive; and
- QUEST3: Quick Scan.

In general, the tools will assist in identifying the program characteristics and policies and procedures that contribute to a quality experience for children aged six to 12 participating in recreation and sports programs.

The major emphasis of this report is an investigation of the reliability and validity of QUEST2, followed by an exploration of the content validity of QUEST1 and the creation of QUEST3. The next section of this chapter outlines the methodology used to assess QUEST2.

2.1 Observing the Child's Experience - Comprehensive (QUEST2)

QUEST2 is an observational tool intended to be used by trained raters. It was developed to measure different aspects of a child's experience while participating in a recreation or sport program. It is organized into sections that reflect the **Guiding Principles and Objectives for Healthy Child Development in Recreation Settings**⁶. Specifically, these sections include:

- leader/child interactions;
- supervision/safety;
- child/peer interactions;
- leader behaviour and interactions;
- program/activities;
- environment/facility; and
- equipment/materials.

Each of the sections contain a number of items that a rater can view in a recreation/sport program. The raters would observe the program and rate it on each of the items. A definition is provided for each item as well as descriptions of sample behaviours that the rater should look for in scoring the item.

A copy of the version of QUEST2 used for this study is provided in Appendix A (under separate cover).

6

Please see, "Final Report, Phase I, The Quality Assurance Standards Project. Parks and Recreation Ontario, 1996" for details related to the tools' development.

2.2 Overview of the Approach Used for the Reliability/Validity Study of QUEST2

The assessment of the reliability and validity of QUEST2 was composed of three generic phases:

internal consistency and inter-rater reliability assessment;
assessment of the tool's construct validity; and
assessment of the tool's content validity.

2.2.1 Internal Consistency and Inter-Rater Reliability of QUEST2

In general, *internal consistency reliability* measures the degree that the scores on one item in a scale or section of QUEST2 are consistent with the scores for other items in the same section. In other words, it is a measure of the inter-correlations among items in a section. If the items in a section are intending to measure the same "thing" (e.g., leader/child interactions) then it is reasonable to assume that a program which scores highly on one item will score highly on other similar items. In other words, items should be correlated with each other and the scale ought to be internally consistent.

There are a number of statistics and methods available to measure internal consistency reliability. One of the most widely used is coefficient Alpha⁷. It was the statistic of choice for this study.

Inter-rater reliability measures the degree of agreement and consistency of agreement among different raters who score the same program. Again, a number of methods are available to measure inter-rater reliability. The methods used for this study are the intraclass correlation and the Spearman-Brown prophecy formula adjusted for the number of raters⁸.

The following tasks were undertaken with respect to the reliability assessment.

2.2.2 Orientation

This step included meetings and discussions regarding the approach and a brief literature review related to reliability studies of similar tools.

⁷ Cronbach, L. J. (1951), "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16, 297-334.

⁸ Shrout, P. E. & Fleiss, J. L. (1979), "Intraclass Correlations: Uses in Assessing Rater Reliability," *Psych. Bulletin*, Vol. 86, No. 2, 420-428.

MacLennan, R. N. (1993), "Interrater Reliability with SPSS for Windows 5.0," *The American Statistician*, Vol. 47, No. 4, 292-296.

2.2.3 First Pilot Test

Prior to administration of the full assessment, two pilot tests were undertaken. The first pilot test consisted of the following tasks:

- selection and training of the raters;
- administration of the ratings for the pilot test;
- data entry and analysis of the pilot ratings;
- meeting with raters to discuss the results; and
- modifications to QUEST2.

Two female raters were hired and trained in the use of QUEST2. Eight sites were selected and rated. They included:

- an instructional swimming program;
- a jujitsu program;
- an instructional hockey skills program;
- a junior basketball program;
- a discovery camp program;
- a nature camp program;
- an adventure camp program; and
- a municipal camp program.

An item analysis was done on the ratings, as well as the calculation of very preliminary measures of internal consistency.

The analysis also identified those items for which the raters did not have close agreement. Once these were identified, a meeting was convened with the raters to better understand the variance in their ratings. In this way, we could better isolate the deficiencies in the tool that caused the variations in ratings and make appropriate modifications to the tool and its instructions.

The above analysis and consultations resulted in modifications to QUEST2. Without detailing each of the scores of modifications that were done, they fell into a number of generic categories:

- removal or rewording of terms, observations and concepts whose interpretation was *obviously and highly* subjective in nature;
- creation of a definition or focus for each item;
- changes to the training curriculum/instructions for raters;
- removal and replacement of specific items;

changing the wording and rating procedures to minimize item ratings that were "not applicable";

movement of items to more appropriate sections of the tool based on their content similarity; and

changing the assessment design to include both a male and a female rater, given the intended users of QUEST2 will be both males and females.

Content feedback was also solicited and received from several known experts in child development and others. These informed personnel reviewed the tool to ensure that the concepts were still intact and comprehensive. Their comments resulted in further refinements.

Once the modifications were complete, the next draft of QUEST2 was used for the second pilot test.

2.2.4 Second Pilot Test

Similar to the first pilot test, the second test included a selection of the sites to be rated, administration of the ratings, analysis of the results and further modifications to QUEST2.

This test involved a male and female rater scoring four programs:

two swimming instruction programs;
a karate program; and
an art class.

An analysis of the items and inter-rater reliability was again undertaken, as well as an investigation into the variances in the ratings, as detailed above. This resulted in further modifications to the instrument in preparation for the full assessment.

2.2.5 Design of the Reliability Assessment

This step involved the design of the sample of sites to be rated for the full assessment. In designing the sample, the following factors were considered:

the size of the sample;

the need for QUEST2 to be assessed across the different types of programs for which it was intended. These were:

Specialized Instructional Programs;
After School/General Interest Programs;
Community Sports Programs;
Fine Arts Programs; and
Instructional Sports Programs.

the need for the tool to be used in programs delivered by private, non-profit and municipal operators; and

the need to ensure that the tool was assessed across programs in both urban and rural settings.

The table below details the sample design that was developed based on consultation with Parks and Recreation Ontario and with consideration for the resources available to the project.

The table indicates that 36 sites were visited across each of the program types.

The design ensured that the testing process was conducted across a broad range of settings representative of the spectrum of sport and recreation services available for school age children, and also representative of a cross-section of urban and rural settings (and diverse socio-economic conditions, where possible).

The testing was conducted in south-central Ontario, including sites in downtown Toronto, major suburban sites such as Etobicoke and Mississauga, mid-sized urban areas such as Oakville, Hamilton, Burlington and Guelph, and small rural sites. Service providers included municipalities, YMCA's, community based non-profit groups, sports clubs and private operators.

Sample Design for Assessment of QUEST2

Type of Program	Total	Locations	Service Providers
Specialized Instructional Programs	4	Mississauga (2), Oakville, North York	City of Mississauga City of North York Private Operator
After School/ General Interest Programs	5	Hamilton, Oakville, East Hamilton, Georgetown, Etobicoke	City of Hamilton Town of Halton Hills City of Etobicoke
Community Sports Programs	8	Georgetown (2), Toronto (2), Oakville, North York, Mississauga (2)	Town of Halton Hills City of Toronto Town of Oakville Community Sports Club YMCA City of Mississauga
Fine Arts	7	East Hamilton, City of Toronto (2) Oakville (2), City of Guelph	City of Hamilton City of Toronto Town of Oakville Non-Profit Community Group YMCA
Instructional Sports Programs	12	Milton (2), Guelph (3), Mississauga, Georgetown (2), East York (2), North York, Oakville	Town of Milton City of Guelph City of Mississauga Borough of East York City of North York YMCA
Total	36		

In preparation for the data collection, the Ministry of Citizenship and Culture provided names of potential sites that could be contacted for program observation. These contacts were incorporated into the site list compiled by the project team.

2.3 Construct Validity

One central purpose of QUEST2 is to take constructs like leader/child interaction, supervision, and child/peer interactions, quantify them on scales, and use numbers to measure the “amount” of the construct that a children’s recreation program possesses.

The purpose of construct related evidence of validity is to determine the degree that the scores on QUEST2 are a true reflection of the construct. This validity can be determined by relating the scores of QUEST2 with other empirical evidence. This means computing correlations with other measures that are accepted to reflect quality in children’ recreation programs (e.g., ratio of leaders to children and leader qualifications). We would at least like to see that the correlations were in the right direction, based on the theory underlying QUEST2.

With respect to the construct validity of QUEST2, the following external measures were identified:

leader/child ratios; and
leader education.

These data were collected for each program rated in the full reliability assessment.

2.3.1 Content Validity

Theoretically, thousands or even millions of items could have been developed for QUEST2 to measure the same constructs. Content related validity is defined as the degree that the items in this particular instrument are representative of the entire domain of items that could measure the construct. In other words, do these particular items appropriately represent all aspects of the construct that could be measured?

For a number of reasons, true content-related validity is hard to measure, mainly due to the fact that it is virtually impossible to describe all the possible items that might be included in a construct.

However, an approach that has been used is to obtain “experts” ratings of the degree that the items reflect their opinion of the construct. In other words, how important do "experts" think each item is as a factor impacting on the healthy development and well being of children. We would hope to see strong agreement across a number of experts that it was, in fact, important.

Thus, the tasks required for content validity assessment were:

identifying "experts" for the test of content validity;

designing the content validity package and questionnaire;
administering the content validity assessment; and
data entry of expert ratings and analysis of results.

The following people were identified, contacted and agreed to participate in the content validity process.

Personnel Involved in QUEST2 Content Validity Process

Mary Gordon	Parenting Program Office, Toronto Board of Education
Dr. Daniel Keating	Department of Human Development, OISE
Lauri Landy	Faculty of Early Childhood Education, Seneca College
Dr. Donna Lero	Department of Family Studies, Guelph University
Judith Levcoe	Vice-President, Association Programs, YMCA of Greater Toronto
Dr. Dan Offord	Director, Centre for Studies of Children at Risk, McMaster University
June Pollard	Chairperson of the School of Early Childhood Education, Ryerson
Beth Stockton	Supervisor, Waterpark Place Child Care Centre, George Brown College
Dr. Rick Volpe	Professor of Human Development, University of Toronto

The package was e-mailed to the experts on November 21, 1997, with a follow-up "reminder" e-mail on December 5, and a telephone follow-up on December 10, 1997.

2.4 The Review of Best Practices (QUEST1)

QUEST1 is a checklist that will be used by managers and staff of recreation and sports programs. It is intended to identify the existence of different structures and practices in recreation and sport programs that could impact on quality. It is also organized into sections that reflect the **Guiding Principles and Objectives for Healthy Child Development in Recreation Settings**. Specifically, these include:

staff/child interactions;
supervision/safety;
program/activities; and
environment/facility.

Each of the sections contain a checklist of structures and practices that could impact on quality in the delivery of a recreation/sport program. A copy of the most recent version of QUEST1 is provided in Appendix B (under separate cover).

The checklist is to be completed separately by a manager and staff person. In practice, *"...the manager, should have copies of all policies and procedures on hand to check whether or not they are complete. The manager should also have a good working knowledge of day-to-day practices, learned by observation over time. The staff person conducting the review should rely upon his or her own working knowledge of policies and day-to-*

Working Draft Document - For Discussion Purposes Only - Not to be Reproduced in Any Form Without Permission of PRO
day practices without referring to training or policy and procedures manuals....Once both parties have completed the review, the manager should compare the results and utilize the Standards Review Follow-up Form to assist with developing strategies for improvements¹⁹

Because the Quality Assurance Standards Survey is a checklist, it is not quantifying constructs on scales and using "scores" or ratings to measure the "amount" of the construct that a recreation program possesses, as QUEST2 does. Its intention is to show program personnel where deficiencies may exist using the checklist, and to stimulate dialogue and appropriate modifications.

For this reason, a reliability assessment of QUEST1 was not undertaken. However, a study of the content validity of QUEST1 was administered.

2.4.1 Overview of the Validity Assessment Process of QUEST1

In a manner similar to QUEST2, the content validity of QUEST1 was measured by sending the checklist to a sample of experienced recreation service providers in Ontario. Each provider was asked to rate the items in the checklist with respect to how important they were as quality assurance factors in recreation and sports programs. The same personnel were also asked to identify any gaps or deficiencies in the checklist. The table below presents the people who agreed to assist with this process.

Personnel Involved in QUEST1 Content Validity Process

Joy Thompson	Manager of Recreation Services, Town of Halton Hills
Jennifer Reynolds	Director of Leisure Services, Town of Milton
Susan Burt	Manager of Recreation, City of Mississauga
Rick Brown	Director of Recreation, Town of Woodstock
Sue Mundick	Director of Recreation, North East District, City of North York
Janet Ellis	Director of Recreation, Borough of East York
Louise Veres	Manager of Recreation, City of St. Catherines
Linda Cork	Manager of Recreation, Town of Ajax Parks and Recreation Department
Lorene Bodiam	City of Etobicoke Parks and Recreation Department
Darlene Joslin	Director of Recreation, Town of Richmond Hill
Linda Cottés	YMCA of Metro Toronto Children's Services

After final modifications to QUEST1, the validity study package was e-mailed on December 9, 1997.

9

Quality Assurance Standards Survey.

3.0 Analysis of the Scales: Observing the Child’s Experience (QUEST2)

3.1 Description of QUEST2 Scales

The version of QUEST2 used for the initial reliability and validity assessment was composed of seven categories or scales. The table below presents the category descriptions, their focus and the items in each scale.

Description of the Categories/Scales of QUEST2

Categories/Scales	Description	Items
Leader/Child Interactions (Seven Items)	This section focuses on the interactions between leaders and children and the impact these interactions have on healthy child development and the development of positive behaviour patterns.	Warmth Interest Respect Individualized Approach Involvement Positive Leadership Children are Given Priority
Supervision/Safety (Four Items)	This section focuses on the manner in which children are supervised to ensure safety at all times.	Awareness Age/Stage Appropriate Activity Risk Awareness Child Supervision in Transition Areas
Child/Peer Interactions (Four Items)	This section looks at the extent to which the children in the program interact with each other in a positive manner.	Familiarity Among Children Respect and Cooperation Among Children Inclusionary Behaviour Atmosphere
Leader Behaviour and Interactions (Three Items)	This category examines whether the Interactions between leaders are appropriate and reflect positive role modelling for children.	Appropriate Behaviour and Language Discretion with Confidential Matters Team Effort
Program/Activities (Four Items)	The section examines whether the activities are well-planned and implemented in such a way as to support the healthy development of the children participating.	Program Planning Balance, Variety and Choice in Activities Pace of Activities Individual Growth Through Group Development
Environment/Facilities (Five Items)	This category focuses on whether the environment/facility is appropriate for the type of activity/program and whether it has all the elements necessary to support healthy child development.	Maximizing Use of Available Facility Amenities Activity Appropriate Space Welcoming Atmosphere Physically Comfortable Space Areas Leading to the Program Area Site Safety
Equipment/Materials (Three Items)	This section looks at whether the equipment and materials are appropriate and support the activity.	Developmentally Appropriate Quantity Culturally Appropriate Materials
Total 31 Items	Total seven sections (scales)	

3.2 Results of the Initial Reliability and Validity Study of QUEST2

This section is organized as follows:

results of the internal consistency reliability study of QUEST2 scales;
results of the Interrater reliability study;
results of the construct validity study; and
results of the content validity study.

3.3 Internal Consistency Reliability

In the pages that follow, the following information is presents for each QUEST2 scale:

inter-item correlations;
scale/item total correlation;
corrected squared multiple correlation;
alpha if the item deleted; and
alpha for the scale.

3.3.1 The Leader/Child Interactions Scale

The table below presents the item statistics for the leader/child interactions scale as discovered for the 36 sites that were rated.

Inter-Item Correlations - Leader/Child Interactions

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.7601	.6462	.8449	.1986	1.3073	.0035

The items in this scale have an average correlation of .76, ranging from a low of .65 to a high of .84.

The table below presents the item-total statistics, as well as the calculation of alphas for each item and the scale.

Item/Total Correlation and Alphas - Leader/Child Interactions Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Warmth	.8412	.7574	.9497
Interest	.9096	.8450	.9447
Respect	.8248	.7616	.9510
Individualized Approach	.7809	.6316	.9543
Involvement	.8351	.7468	.9505
Positive Leadership	.8909	.8217	.9461
Children are Given Priority	.8691	.7818	.9475
Scale Alpha	.9561		

The overall alpha for the scale is .96, indicating excellent reliability. In addition, all items in the scale contribute almost equally to the scale's alpha.

3.3.2 The Supervision/Safety Scale

The item statistics for the supervision/safety scale are presented in the table below.

Inter-Item Correlations - Supervision/Safety

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.4412	.2410	.6960	.4550	2.8884	.0304

The items in this scale have an average correlation of .44, ranging from a low of .24 to a high of .70.

As before, the table below presents the item-total statistics, as well as the calculation of alphas for each item and the scale.

Item/Total Correlation and Alphas - Supervision/Safety Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Awareness	.6096	.4302	.6398
Age/Stage Appropriate	.6766	.6096	.6046
Activity Risk Awareness	.5606	.4870	.6698
Child Supervision in Transition Areas	.3538	.1501	.8068
Scale Alpha	.7406		

The alpha for this scale is .74, indicating good reliability. However, the item "child supervision in transition areas" had the lowest correlations with the other items in the scale and the total score for the scale. You can also see in the table above that if this item is removed from the scale, the scale alpha rises to .81.

3.3.3 The Child/Peer Interactions Scale

The item statistics for the child/peer interactions scale are presented below.

Inter-Item Correlations - Child/Peer Interactions

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.4531	.2471	.6280	.3809	2.5413	.0170

The items in this scale have an average correlation of .45, ranging from a low of .25 to a high of .63.

The table below presents the item-total statistics, as well as the calculation of alphas for each item and the scale.

Item/Total Correlation and Alphas - Child/Peer Interaction Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Familiarity Among Children	.6151	.4475	.6793
Respect and Cooperation Among Children	.5820	.4643	.6987
Inclusionary Behaviour	.5563	.4832	.7165
Atmosphere	.5186	.3922	.7349
Scale Alpha	.7639		

The alphas for the child/peer interactions scale is .76 indicating good reliability. In addition, little difference was found with respect to the contribution each item makes to the scale alpha.

3.3.4 The Leader Behaviour and Interactions Scale

As a reminder, this scale consisted of three items, appropriate behaviour and language, discretion with confidential matters and team effort.

Across the 36 sites, discretion with confidential matters was consistently rated as excellent, providing no discrimination (or variance) in scores among the sites. In addition, if there was only one leader, the raters had difficulty scoring team effort, and appropriately left this item unscored.

Thus, the statistics presented below are for the 44 ratings which had multiple leaders and do not include the item, "discretion with confidential matters".

Inter-Item Correlations - Leader Behaviour and Interactions

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.7282	.7282	.7282	.0000	1.0000	.0000

The two items remaining in this scale have an average correlation of .73.

The table below presents the item-total statistics, as well as the calculation of alphas for each item and the scale.

Item/Total Correlation and Alphas - Leader Behaviour and Interactions Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation
Appropriate Behaviour and Language	.7282	.5303
Team Effort	.7282	.5303
Scale Alpha	.8209	

The alphas for the scale formed by the two items under leader behaviour and interactions is .8209, indicating good reliability.

3.3.5 The Program/Activities Scale

The item statistics for the program/activities scale are presented below.

Inter-Item Correlations - Program/Activities

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.5261	.4327	.6102	.1776	1.4104	.0039

The items in this scale have an average correlation of .53, ranging from a low of .43 to a high of .61.

The item-total statistics, as well as the calculation of alphas for each item and the scale are presented below.

Item/Total Correlation and Alphas - Program/Activities

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Program Planning	.6533	.4425	.7616
Balance, variety and Choice in Activities	.6668	.4534	.7552
Pace of Activities	.6616	.4660	.7573
Individual growth Through Group Development	.5690	.3449	.7993
Scale Alpha	.8164		

The alpha for the child/peer interactions scale is .82 indicating good reliability, and the removal of any of the items would lower the alpha for the scale.

3.3.6 The Environment/Facility Scale

The item statistics for the environment/facility scale are presented in the following table.

Inter-Item Correlations - Environment/Facility

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.3054	.0100	.6220	.6321	62.0074	.0337

The items in this scale have an average correlation of .31, ranging from a low of .01 to a high of .62.

The item-total statistics, as well as the calculation of alphas for each item and the scale are presented below.

Item/Total Correlation and Alphas - Environment/Facilities Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Maximizing Use of Available Facility Amenities	.5728	.3540	.6439
Activity Appropriate Space	.5661	.4484	.6368
Welcoming Atmosphere	.5681	.5061	.6417
Physically Comfortable Space	.3793	.2981	.6981
Areas Leading to the Program Area	.2670	.4400	.7301
Site Safety	.3891	.2457	.6948
Scale Alpha	.7142		

The alpha for the environment/facility scale is .71 indicating adequate/good reliability.

Again, however, the reader will notice that removing the item, "areas leading to the program" will boost the alpha of the scale to .73. Chapter four will address this issue in more detail.

3.3.7 The Equipment/Materials Scale

The item statistics for the equipment/materials scale are presented below.

Inter-Item Correlations - Equipment/Materials

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-item Correlations	.4846	.2610	.9017	.6407	3.4543	.1046

The items in this scale have an average correlation of .48, ranging from a low of .26 to a high of .90.

The item-total statistics, as well as the calculation of alphas for each item and the scale are presented below.

Item/Total Correlation and Alphas - Equipment/Materials Scale

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Developmentally Appropriate	.8191	.8164	.3877
Quantity	.7906	.8131	.4269
Culturally Appropriate Materials	.2828	.0847	.9481
Scale Alpha	.7638		

The alpha for the equipment/materials scale is .76 indicating good reliability.

You will note again that if the item, "culturally appropriate materials" is removed from the scale the alpha rises to .95.

Given the nature of certain programs in sport and recreation, culturally appropriate materials does not always apply (e.g., a hockey practice). Because of this, the ratings for this item were completed as not applicable or were rated in a manner that was not consistent with the other items in the scale. In fact, this item had an average correlation of only .28 with the other items in the scale, and with the scale score. Further examination also showed that the ratings were missing for almost 50% of the sports related program types. Chapter four contains further discussion regarding the handling of this item.

3.3.8 The Total Score

The item statistics for the total score are presented below.

Inter-Total Correlations- Scale Totals

Description	Mean	Minimum	Maximum	Range	Max/Min	Variance
Inter-total Correlations	.7423	.5171	.8784	.3613	1.6987	.0078

The scale totals in QUEST2 have an average correlation of .77, ranging from a low of .63 to a high of .86.

The scale total/total statistics, as well as the calculation of alphas for each item and the scale are presented below.

Sub-Total/Total Correlation and Alphas - Total Score

Item	Scale Item Total Correlation	Corrected Squared Multiple Correlation	Alpha If Item Deleted
Leader/Child Interactions	.8462	.7806	.9339
Supervision/Safety	.9105	.8672	.9069
Child/Peer Interactions	.8434	.7644	.9152
Leader Behaviour and Interactions	.8324	.8287	.9307
Program Activities	.8444	.8301	.9151
Environment/Facility	.8484	.8959	.9132
Equipment/Materials	.7463	.8270	.9262
Total Alpha	.9308		

The alpha for the total is .93 indicating excellent reliability.

3.4 Inter-Rater Reliability of QUEST2

As mentioned, the statistic used to measure the inter-rater reliability of QUEST2 was the intraclass correlation. This correlation is widely used to measure the relationship between two independent raters who rate a group of sites or objects; the higher the value of this measure, the better the inter-rater reliability. The table below presents the intraclass correlation calculated for each of the scales of QUEST2, and for the total score.

Intraclass Correlation for QUEST2 Scales and Total Score

Scale/Total	Intraclass Correlation (ICC 3,1) ⁵
Leader/Child Interactions	0.7832
Supervision/Safety	0.8243
Child/Peer Interactions	0.6473
Leader Behaviour and Interactions	0.9450
Program Activities	0.6491
Environment/Facility	0.2781
Equipment/Materials	0.1408
Total Score ¹⁰	0.8184

The intraclass correlations show good to excellent inter-rater reliability for leader/child interactions, supervision/safety, leader behaviour and interactions and total score. The correlations are barely adequate for child/peer interactions and program/activities. They are not adequate for environment/facility and equipment/materials. Chapter four addresses this issue in more detail.

10

With the scale equipment and materials removed.

3.5 Construct Validity of QUEST2

As stated earlier in this report, construct validity can be determined by relating the scores of QUEST2 with other empirical evidence. This means computing correlations with other measures that are generally accepted to reflect quality in children's recreation programs. QUEST2 total and scale scores were correlated with the following external measures:

leader/child ratios; and
leader education.

3.5.1 Leader/Child Ratios

The leader/child ratios were calculated by dividing the total number of children in the program by the number of leaders; the larger the number of leaders relative to children the smaller this number.

We would hope that higher scores on QUEST2 would be associated with smaller ratios (a negative correlation). The table below presents these results for QUEST2 scales and total score.

Correlation of Staff/Child Ratios to QUEST2 Scales and Total Scores

Measure	Leader/Child Interactions	Supervision/Safety	Child/Peer Interactions	Leader Behaviour and Interactions	Program Activities	Environment/Facility	Equipment / Materials	Total Score
Staff/Child Ratios	-.2557	-.4639*	-.2286	-.4004*	-.2473	-.2404	.0488	-.3822*

The scales and total scores have moderate and in some cases significant¹¹ negative correlations ("*") with staff to child ratios.

Although the correlations range from modest to significant, they do indicate a relationship that points in the right direction for QUEST2.

3.5.2 Years of Education of Leaders

Raters recorded the highest education attained by program leaders. This information was translated into the number of years of education of the leaders. For example, if the highest education a leader attained was Grade 11, he/she had received 12 years of education. If a leader had completed high school and a four year university degree, he/she had attained 18 years of education.

¹¹ At the 95% level of confidence.

In only one case did a leader have a diploma in child development, although four reported taking child development courses as part of their university degree curriculum.

We would expect that the scale and total scores for QUEST2 would have positive correlations with years of education. In other words, the greater the number of years of education of program leaders, the higher the scores on QUEST2. The table below presents the results.

Correlation of Years of Education to QUEST2 Scales and Total Scores

Measure	Leader/Child Interactions	Supervision/Safety	Child/Peer Interactions	Leader Behaviour and Interactions	Program Activities	Environment/Facility	Equipment / Materials	Total Score
Years of Education	.2906	.3809*	.2305	.6550*	.2620	.2887	.1457	.5463*

Again, we found moderate and in some cases significant correlations between years of education and the scores on QUEST2. As before, the correlations are modest but certainly in the right direction.

Given that both measures of the construct validity of QUEST2 revealed relationships that were in the "right" direction, and considering the range of programs and the size of the sample, these results indicate that QUEST2 has a degree of construct validity to support its use.

3.6 Content Validity of QUEST2

At the time of writing this report, nine experts in child development had completed and returned their forms. The table below presents the experts' average ratings of each of the items of QUEST2 on a scale from one to five, where one was not at all important as a factor effecting the healthy development of children and five was extremely important.

Average Rating of QUEST2 Items and Scales by Child Development Experts

Scale/Item	Average Rating	Scale/Item	Average Rating
Leader/Child Interactions Scale	4.57	Leader Behaviour and Interactions Scale	4.44
Warmth	4.89	Appropriate Behaviour and Language	4.56
Interest	4.78	Discretion with Confidential Matters	4.67
Respect	4.89	Team Effort	4.11
Individualized Approach	4.56		
Involvement	4.11	Program/Activities Scale	4.33
Positive Leadership	4.33	Program Planning	4.44
Children are Given Priority	4.44	Balance, Variety and Choice in Activities	4.44
		Pace of Activities	4.22
Supervision Safety Scale	4.47	Individual Growth Through Group Development	4.22
Awareness	4.78		
Age/Stage Appropriate	4.56	Environment Facilities Scale	3.91
Activity Risk Awareness	4.56	Maximizing Use of Available Facility Amenities	3.89
Child Supervision in Transition Areas	4.00	Activity Appropriate Space	4.00
		Welcoming Atmosphere	4.33
Child/Peer Interactions Scale	4.47	Physically Comfortable Space	3.56
Familiarity Among Children	4.56	Areas Leading to the Program Area	3.22
Respect and Cooperation Among Children	4.56	Site Safety	4.44
Inclusionary Behaviour	4.44		
Atmosphere	4.33	Equipment/Materials Scale	4.30
		Developmentally Appropriate	4.44
		Quantity	3.89
		Culturally Appropriate Materials	4.56
All Items 4.36			

Twenty seven of the thirty one items (87%) had an average rating of 4 or higher. The most highly rated items were warmth and respect, both rated at 4.89. The lowest items were Areas Leading to the Program Area (3.22), Physically Comfortable Space (3.56), Maximizing Use of Available Facilities (3.89) and Quantity (3.89). The average overall rating for all items was 4.36. According to the opinions of the experts in child development, QUEST2 has excellent content validity.

The table below list the average scores for the scales themselves, ordered from highest to lowest.

Average Ratings of Scales by Child Development Experts

Scale/Item	Average Rating
Leader/Child Interactions Scale	4.57
Supervision Safety Scale	4.47
Child/Peer Interactions Scale	4.47
Leader Behaviour and Interactions Scale	4.44
Program/Activities Scale	4.33
Equipment/Materials Scale	4.30
Environment Facilities Scale	3.91

The scale rated the highest was leader/child interactions. On the other hand, the scale rated the lowest was environment/facilities.

3.7 Summary of Initial Findings QUEST2

The table below presents a summary of the findings of the initial reliability/validity assessment of QUEST2.

Summary of Reliability/Validity Assessment of QUEST2

Categories/Scales	Internal Consistency Reliability	Intraclass Correlation (ICC 3,1)	Content Validity ¹²	Construct Validity	
	Alpha			Staff/Child Ratio	Education
Leader/Child Interactions	0.96	0.78	4.57	-0.26	0.29
Supervision/Safety	0.74	0.82	4.47	-0.45	0.38
Child/Peer Interactions	0.76	0.65	4.47	-0.23	0.23
Leader Behaviour and Interactions	0.82	0.95	4.44	-0.40	0.66
Program/Activities	0.82	0.65	4.33	-0.25	0.26
Environment/Facilities	0.71	0.28	3.91	-0.24	0.29
Equipment/Materials	0.76	0.14	4.30	0.05	0.15
Total	0.93	0.82	4.36	-0.38	0.54

12

Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

3.8 Content Validity of QUEST1

Eight experienced recreation service providers completed and returned their forms. The providers again used a scale from one to five, where one was not at all important (as a factor effecting the healthy development of children) and five was extremely important. A detailed listing of the ratings for all items is presented in Appendix C.

With respect to the individual items, eighty eight percent of the items were rated four or higher, and 53% were rated 4.5 or higher. The table below presents the average ratings of the providers for each *section* of QUEST1.

Recreation Providers' Average Ratings of QUEST1 Sections

Sections	Average Rating
Leader Hiring/ Recruitment/ Placement Procedures	4.50
Leader Behaviour	4.38
Leader Cooperation	4.07
Leader Orientation	4.54
Emergency Preparedness	4.80
Child Abuse Prevention, Detection and Reporting	4.79
Injury/Illness Preparedness and Prevention	4.54
Child Supervision	4.81
Record Keeping	4.69
Connections to Home	4.50
Leadership Qualification/ Assessment	4.33
Sanitation	4.13
Preparation and/or Serving of Food	3.96
Connections to Home	3.83
Program Evaluation Process	4.25
Program Planning	4.00
Program Transitions	4.00
Environment/Facility	4.88
All Sections	4.32

Eighty nine percent of the sections (16 out of 18) had an average rating of four or higher. In fact, half of the sections were rated 4.5 or higher. Overall, the providers rated all sections at an average of 4.33. It is clear that, according to the recreation providers, QUEST1 has excellent

content validity.

3.9 Preliminary Conclusions/Discussion Regarding QUEST2

Based on the findings to date, we can make the following preliminary conclusions regarding the initial version of QUEST2.

The following scales appear to have sufficient internal consistency reliability, interrater reliability, construct and content validity to support their use as measures of factors effecting the healthy development and well-being of children in sport/recreation programs:

leader/child interactions;
supervision/safety; and
leader behaviour and interactions.

The following scales appear to have sufficient internal consistency reliability, as well as construct and content validity. However, their interrater reliability is just adequate.

child/peer interactions; and
program/activities.

The following scales appear to have sufficient internal consistency reliability. However, their lower interrater reliability, construct and content validity implies the need for further modifications and research prior to supporting their use as measures of factors effecting the healthy development and well-being of children in sport/recreation programs.

environment/facilities; and
equipment/materials.

3.9.1 Discussion

The initial study findings highlight the principal challenge facing QUEST2 - the need and the difficulty obtaining reliable *observational* measurements across a broad spectrum of sports and recreation programs.

3.9.2 Specific Items

One item (teamwork in the leader behaviour and interactions section), although important, does not appear to be observable in every sports/recreation program.

In addition, three items were quite consistently rated in a different fashion from the other items in their sections. These were:

child supervision in transition areas in the supervision/safety section;
areas leading to the program area in the environment/facility section; and

culturally appropriate materials in the equipment/materials section.

3.9.3 Interrater Reliability

The study has shown that the initial version of the tool was somewhat less than successful with respect to its interrater reliability. We feel this was due to the following:

All sport and recreation programs do not have the same structural or delivery characteristics, and certain items in the current version of the tool are more appropriate for and/or observable in some programs and not others. In addition, the raters are observing and making judgements about a reasonably complex set of behaviour patterns and environments - making their adjudication more difficult.

The raters' scoring of QUEST2 can be influenced by the different knowledge they have about the requirements of a program *that are specific to the program*. For example, under environment/facility and equipment/materials, different raters may have a different level of knowledge regarding the "appropriate" environment, equipment and space requirements for a karate program or a gymnastics program - what one rater considers appropriate, the other may not.

The raters' scoring of QUEST2 obviously can be influenced by their backgrounds, their attitudes and the personal perceptions they "bring" to the ratings. For example, the Guiding Principles and Objectives for Healthy Child Development In Recreation Settings imply a new "position" with respect to how children should be treated in sports and recreation programs. This "position" and the underpinning principles on which it is based need to be inculcated into raters prior to their use of the tool.

3.9.4 The Length of the Tool

Administration of the tool can be a lengthy process. The raters reported that the time required to complete the tool was between 45 minutes to an hour. If detailed notes were taken, the administration was even longer. Numerous other reviewers and practitioners have also reported that the tool is too lengthy. The length of the tool could effect the degree that it will be used widely and consistently across the province, especially given the voluntary compliance nature of the High Five Program.

It is our view that a shorter version is needed as well. This tool would contain a selection of items and scales that have demonstrated the strongest psychometric properties and yet still adequately reflect the Guiding Principles and Objectives for Healthy Child Development In Recreation Settings.

The items and scales that have shown less than acceptable psychometric characteristics could be removed from QUEST2 and be incorporated into QUEST1 and/or the appropriate sections of the policy manual to which they relate.

3.10 Conclusions

The above discussion highlights the steps that need to be taken to improve the tool.

Further analysis of the results of the reliability and validity testing with a view to enhancing the psychometric properties of the tool;

Identification of the appropriate and always observable items that exist in every sport/recreation program. Include these in the main body of the tool and the others as supplementary items;

Creating a shorter version of the tool;

Using the results of this study in the design of the training curriculum/method for using the tool; and

Introducing the tool and its use in a measured and planned way, and consideration of continued testing of the tool during the High Five Program pilot phase. This testing could also create the initial database for the development of norms that would eventually allow programs to view their results relative to a "provincial standard".

The next chapter of the report addresses the above.

4.0 Additional Analyses and QUEST2 Modifications

Based on the initial findings, and through detailed discussions and meetings with the raters and High Five personnel, a number of modifications were made to the content of and observational approach used with QUEST2. In general, the decisions were based on the ratings of the experts, the item's psychometric properties and the opinions of the High Five team directing the project.

- 1 The following items were dropped from the environment/facilities and equipment/materials scales:

maximizing use of available facilities and amenities (lower importance according to experts);
areas leading to the program area (poorer psychometric qualities, lower importance according to experts);
physically comfortable space (lower importance according to experts) and;
culturally appropriate materials (difficulty observing in most sport/recreation programs, poor psychometric qualities). Given the importance of the item according to the experts, it will be incorporated in the policy manual and QUEST1.

- 2 The item "site safety" was moved to the safety/supervision section.

- 3 The item "Team Effort" was repositioned as a supplementary item in the leader behaviour and interactions section - to be completed when more than one leader is directing the program.

- 4 The following items were combined into a new scale titled "Program Characteristics and Supports":

Program Planning
Activity Appropriate Space
Welcoming Atmosphere
Developmentally Appropriate
Quantity of Materials/equipment
Balance, Variety and Choice in Activities
Pace of Activities
Individual Growth Through Group Development

Overall, two "new" scales were created - safety/supervision (addition of site safety) and program characteristics/supports. The table on the next page details the redefined structure of QUEST2.

Redefined Structure of QUEST2

Scale/Item	Scale/Item
Leader/Child Interactions Scale	Leader Behaviour and Interactions Scale
Warmth	Appropriate Behaviour and Language
Interest	Discretion with Confidential Matters
Respect	Team Effort
Individualized Approach	
Involvement	Program Characteristics and Supports
Positive Leadership	Program Planning
Children are Given Priority	Balance, Variety and Choice in Activities
	Pace of Activities
Supervision Safety Scale	Individual Growth Through Group Development
Awareness	Activity Appropriate Space
Age/Stage Appropriate	Welcoming Atmosphere
Activity Risk Awareness	Developmentally Appropriate Equipment/ Materials
Child Supervision in Transition Areas	Quantity of Equipment/Materials
Site Safety	
	Total Items = 27
Child/Peer Interactions Scale	
Familiarity Among Children	
Respect and Cooperation Among Children	
Inclusionary Behaviour	
Atmosphere	

The next two sections present the psychometric characteristics of the two new scales

4.1 The Redefined Safety/Supervision Scale

As before, the table below presents the item-total statistics, as well as the calculation of alphas for each item and the scale.

Item/Total Correlation and Alphas - Safety/Supervision Scale

Item	Scale Item Total Correlation	Alpha If Item Deleted
Awareness	.6699	.6105
Age/Stage Appropriate	.6465	.6224
Activity Risk Awareness	.5821	.6495
Child Supervision in Transition Areas	.3148	.7695
Site Safety	.3157	.7406
Scale Alpha	.7290	

The alpha for this scale (.73) is almost identical to the original scale, indicating good reliability.

4.2 The Program Characteristics/Supports Scale

Item/Total Correlation and Alphas - Program Characteristics and Supports

Item	Scale Item Total Correlation	Alpha If Item Deleted
Program Planning	.5745	.8342
Balance, Variety and Choice in Activities	.6476	.8251
Pace of Activities	.6570	.8237
Individual Growth Through Group Development	.6006	.8318
Activity Appropriate Space	.3114	.8636
Welcoming Atmosphere	.5437	.8379
Developmentally Appropriate	.7130	.8160
Quantity	.6739	.8213
Scale Alpha	.8503	

The alpha for this scale is .85 indicating excellent reliability.

4.3 The Redefined Total Score

The item statistics for the "new" total score are presented below.

Sub-Total/Total Correlation and Alphas - Total Score

Item	Scale Item Total Correlation	Alpha If Item Deleted
Leader/Child Interactions	.8715	.8706
Supervision/Safety	.8622	.8561
Child/Peer Interactions	.8477	.8804
Leader Behaviour and Interactions	.8447	.9098
Program Characteristics/ Supports	.8906	.8485
Total Alpha	.8989	

The alpha for the total is .90 indicating excellent reliability.

4.4 Content and Construct Validity of the Redefined Scales

The table below summarizes both the content (expert ratings) and construct (correlations with staff/child ratio and education of leader) validity of the new scales and the total score using the new scales.

Summary of Validity Assessment of New QUEST2 Scales

Scales	Content Validity ⁵	Construct Validity	
		Staff/Child Ratio	Education
Supervision/Safety	4.47	-0.46	0.46
Program Characteristics/Supports	4.25	-0.24	0.33
Total	4.44	-0.41	0.62

Twenty six of the 27 items in QUEST2 (96%) had an average rating of 4 or higher. The average overall rating for all items was 4.4. According to the opinions of the experts in child development, QUEST2 has excellent content validity.

⁵

Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

For the supervision/safety scale, the experts' average ratings did not change. With respect to the program characteristics and support scale, the experts' average rating is over four at 4.25.

For the supervision/safety scale, the correlation with staff/child ratios did not change. However, the correlation with leader education has increased from .38 to .46 - which is significant at the 95% confidence level.

Considering the program characteristics/support scale, the correlation with staff/child ratio is in the right direction at -.24. In addition, the correlation with education is also in the right direction and significant at .33.

The total score has an average expert rating of 4.4 and correlates significantly and in the correct direction with staff/child ratio and education.

4.5 Interrater Reliability of the Redefined Scales

Preamble

As mentioned, the statistic used to measure the interrater reliability for QUEST2 was the intraclass correlation (ICC). Because of its history in psychometric theory, many of the properties of the ICC are well known. One of the more useful properties is that it allows us to estimate the interrater reliability of our measurement when different numbers of raters view the same program and then pool or average their scores. As stated by Hartmann⁶;

"...under particularly adverse observational conditions, such as might be experienced on a large and crowded playground, or for behaviors that for any reasons are difficult to discriminate, some investigators may find it more efficient to improve reliability by employing multiple observers whose scores will be pooled than to engage in lengthy observer training, an experimental analysis of observer behaviour, or the purchase of costly recording equipment. When such situations occur, ICC (the session reliability for a single pair of observers)...may be used for estimating the interobserver reliability of the pooled observer scores."

In other words, the intraclass correlation as presented earlier in this report is the index of reliability for a single typical rater viewing a program. The reliability index can also be calculated assuming that two or three or any number of raters rate the same program at the same time and pool or average their ratings. This statistic is known as the Spearman-Brown corrected value of the intraclass correlation, adjusted for the total number of raters. In general, the larger the number of raters viewing a program (and averaging their ratings), the higher the value of the adjusted intraclass correlation.

The table below presents the adjusted intraclass correlations for the redefined scales, assuming that one, two and three raters rate the same program at the same time and then pool or average their ratings for the final score.

6

Hartmann, Donald P. Considerations in the Choice of Interobserver Reliability Estimates, Journal of Applied Behavioral Analysis, 1977, (10), 103-116, Number 1 (Spring) 1977.

Intraclass Correlations for Redefined QUEST2 Scales and Total Score

Scale/Total	Intraclass Correlation (ICC 3,1) One Rater	Spearman Brown Prophecy Two Raters	Spearman Brown Prophecy Three Raters
Leader/Child Interactions	.78	.88	.92
Supervision/Safety	.80	.89	.92
Child/Peer Interactions	.65	.79	.85
Leader Behaviour and Interactions	.95	.97	.98
Program Characteristics/Supports	.62	.76	.83
Total Score	.91	.95	.97

The intraclass correlation for supervision/safety has remained virtually the same at .80. The intraclass correlation for program characteristics/supports has moved to .62 from an average of .35 (among the three original scales that it replaced).

In addition, it is clear that the use of two raters substantially improves the reliability. Using two raters, the interrater reliability ranges from good to excellent. Using three raters, the adjusted intraclass correlations are all excellent.

Given the range and complexity of programs in which QUEST2 will be applied, these findings lead us to strongly recommend that during the initial stages of the initiative **the tool be administered by at least two raters and that the final scores for the tool be calculated as the average of their ratings.** If the tool is administered in this way, the interrater reliability indices will range from good to excellent.

4.6 Summary of Findings of the Redefined Version of QUEST2

The table below presents a summary of findings related to the redefined version of QUEST2.

Summary of Reliability/Validity of the Redefined Version of QUEST2

Scale	Internal Consistency Reliability		Spearman Brown Prophecy		Content Validity ⁷	Construct Validity	
	Alpha	Intraclass Correlation (ICC 3,1) One Rater	Two Raters	Three Raters		Expert Ratings	Staff/Child Ratio
Leader/Child Interactions	0.96	.78	.88	.92	4.57	-0.26	0.29
Supervision/Safety	0.73	.80	.89	.92	4.47	-0.46	0.46
Child/Peer Interactions	0.76	.65	.79	.85	4.47	-0.23	0.23
Leader Behaviour and Interactions	0.82	.95	.97	.98	4.44	-0.40	0.66
Program Characteristics/ Supports	0.85	.62	.76	.83	4.25	-0.24	0.33
Total Score	0.90	.91	.95	.97	4.44	-0.41	0.62

4.7 The Shorter Version of QUEST2: QUEST3

Based on the findings of chapter three, it was decided to create a shorter version of QUEST2. This section presents the rationale and construction of this shorter version of QUEST2 - namely QUEST3 (Quick Scan). There were four principal criteria for deciding on the length of QUEST3 and which items should be included in its content.

the need to have a tool that could be administered in 20 to 30 minutes;

the need to insure that the percentage of items in each scale of QUEST3 reflect a similar percentage of items for the same scales in QUEST2;

the item's rating by the experts (content validity); and

the item's psychometric properties.

⁷

Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

4.7.1 Administration Time

The first decision made was with respect to the number of items in QUEST3. QUEST2 required 45 minutes to an hour to administer. The objective for QUEST3 was to have a tool that could be administered in 20 to 30 minutes or in about 40% to 50% of the time it takes to do QUEST2. QUEST2 contains 27 items. Therefore, QUEST3 should contain approximately 10 to 14 items.

In order to facilitate a proper balance with the items in QUEST2, QUEST3 was targeted to contain 14 items. The table below displays the distribution of the items in QUEST3 relative to QUEST2.

Item Distributions QUEST2 and QUEST3

Scale	Number of Items in QUEST2	Percentage of Items in QUEST2	Number of Items in QUEST3
Leader/Child Interactions	7	26%	4
Supervision/Safety	5	19%	2
Child/Peer Interactions	4	15%	2
Leader Behaviour and Interactions	3	11%	2
Program Characteristics/Supports	8	30%	4
Total Score	27	100%	14

4.7.2 The Items Selected for QUEST3

Within each scale, items were chosen that had the highest ratings according to the experts. The table below displays these items.

Structure of QUEST3

Scale/Item	Scale/Item	Scale/Item
Leader/Child Interactions Scale	Child/Peer Interactions Scale	Program Characteristics/Supports
Warmth	Familiarity Among Children	Program Planning
Interest	Respect and Cooperation Among Children	Balance, Variety and Choice in Activities
Respect		Welcoming Atmosphere
Individualized Approach	Leader Behaviour and Interactions Scale	Developmentally Appropriate
	Appropriate Behaviour and Language	
Supervision Safety Scale	Discretion with Confidential Matters	
Awareness		

Activity Risk Awareness ⁵		
--------------------------------------	--	--

4.7.3 Psychometric Properties of QUEST3

The tables below present a summary of the psychometric characteristics of QUEST3 when used as one scale.

Summary of Reliability/Validity of QUEST3

Scale	Internal Consistency Reliability				Content Validity ⁶	Construct Validity		
	Alpha	Intraclass Correlation (ICC 3,1) One Rater	Spearman Brown Prophecy Two Raters	Spearman Brown Prophecy Three Raters	Expert Ratings	Staff/Child Ratio	Education	Correlation with QUEST2 Total Score
QUEST3	0.93	0.88	0.93	0.96	4.60	-0.39	0.46	0.98

As indicated in the table above, QUEST3 possesses excellent psychometric characteristics. The scale produces an alpha of .93 and its intraclass correlation ranges from .88 for one rater to .96 for three. The average ratings by experts is 4.6 (as expected, given the nature of its construction), and its construct validity is acceptable.

In addition, the total score for the QUEST3 was correlated with the total score for QUEST2. The correlation between the two scores is .98, indicating that the QUEST3 total score is an excellent predictor of the QUEST2 total score.

⁵ This item was tied in the expert ratings with "age stage appropriate". This item was chosen as it contributed more to the alpha for the scale.

⁶ Average ratings using a scale from 1 to 5, where 1 is not at all important and 5 is extremely important (as a factor effecting the healthy development of children).

5.0 Scoring QUEST2

This section of the report describes the approach and procedures that can be used to obtain a score for QUEST2. Overall, there are three methods that can be employed with respect to calculating and using the scores.

5.1 Sub-Totals for Each Scale/Category and Overall Total for QUEST2

The first method involves calculating sub-totals for each category or scale of QUEST2. These are computed by simply totalling the ratings given to each item in each section.

In calculating the *overall* total score for QUEST2, "weights" have been computed that can be applied to the sub-totals for each section. These weights were calculated using the experts' ratings of the importance of the items in each category⁵.

The table on the next page presents an example of how the scale sub-totals are calculated, as well as the unweighted and weighted calculations of the sub-totals and total scores.

The ultimate intention of the High Five Program is to have the sub-totals and total scores sent to the High Five office for inclusion in their data base. In this way, provincial norms can be developed that will allow agencies administering QUEST2 to better understand their scores relative to a provincial "standard" for their type of program.

5

The weights were calculated using the Method of Successive Categories. Please see; Shizuhiko Nishisato. Elements of Applied Scaling, An Informal Publication of the Department of Measurement and Evaluation, OISE, 1973, pp 243-257.

Example Scoring Table for QUEST2

Leader/Child Interactions	Example Score	Supervision Safety	Example Score	Child/Peer Interactions	Example Score	Leader Behaviour and Interactions	Example Score	Program Characteristics and Supports	Example Score
Warmth	3	Awareness	2	Familiarity Among Children	6	Appropriate Behaviour and Language	1	Program Planning	2
Interest	2	Age/Stage Appropriate	4	Respect and Cooperation Among Children	5	Discretion with Confidential Matters	4	Balance, Variety and Choice in Activities	3
Respect	5	Activity Risk Awareness	5	Inclusionary Behaviour	4	Team Effort	7	Pace of Activities	4
Individualized Approach	4	Child Supervision in Transition Areas	2	Atmosphere	2	Scale Sub-Total	12	Individual Growth Through Group Development	3
Involvement	1	Site Safety	3	Scale Sub-Total	17			Activity Appropriate Space	2
Positive Leadership	2	Scale Sub-Total	16					Welcoming Atmosphere	4
Children are Given Priority	1							Developmentally Appropriate	4
Scale Sub-Total	18							Quantity	2
								Scale Sub-Total	24
Summary									
Unweighted Scale Sub-Total	18		16		17		12		24
Overall Unweighted Total Score for QUEST2			87						
Scale Weight Factor	1.72		1.44		1.34		1.26		1.00
Weighted Scale Sub-Total	30.96		23.04		22.78		15.12		24
Overall Weighted Total Score for QUEST2			115.9						

The table below provides a general idea of the theoretical range (lowest to highest score) for the sub-totals of each scale in QUEST2 and for the overall total score. This will allow users of the tool to at least see where their score "sits" relative to the lowest and highest score possible and the mid-point.

Theoretical Ranges for QUEST2 Sub-Totals and Total Scores

	Leader Child Interactions	Supervision Safety	Child/Peer Interactions	Leader Behaviour and Interactions	Program Characteristics and Supports
Unweighted Range	7 to 63	5 to 45	4 to 36	3 to 27	8 to 72
Unweighted Mid-Point	35	25	20	15	40
Weighted Range	12.04 to 108.36	7.2 to 64.80	5.36 to 48.24	3.78 to 34.02	8 to 72
Weighted Mid-Point	60.2	36	26.8	18.9	40

QUEST2 Total Score	
Unweighted Range	27 to 243
Unweighted Mid-Point	135
Weighted Range	36.38 to 327.42
Weighted Mid-Point	181.90

The table below provides similar information for the total score of QUEST3.

Theoretical Ranges for QUEST3 Total Score

QUEST3 Total Score	
Range	14 to 126
Mid-Point	70

5.2 Item by Item Scoring

This method of scoring involves reviewing each item in the scales and noting those that have been rated as five or less. Once these items have been identified, one can then move to the appropriate and linked section of the policy manual to better understand the policies and procedures that will help ameliorate the low score. The table below displays the relationship between the items/scales of QUEST2 and the appropriate sections of the policy manual.

5.3 Target Scores for High Five "Accreditation"

As the High Five Program evolves and matures, accreditation scores will also be developed. These would be the scores that programs would need to obtain in order to achieve High Five "accreditation". These scores will be created after the program has a better understanding of the provincial norms and the average ratings of different programs in different sectors, relative to the norms.

5.4 Discussion on the Administration of QUEST3 and QUEST2

Although QUEST3 can be administered in a shorter period of time, it does not contain the number or breadth of items that are in QUEST2. In fact, they are almost cut in half. Given there are fewer items in each section of QUEST3, the alphas for each scale are lower than QUEST2 - and in some cases too low.

QUEST2 was tested for each of its scales, whereas QUEST3 was tested for ALL items combined and its total score only. QUEST3 is, in fact, intended as a quick scan, where its total score will indicate potential weaknesses and the need for more detailed investigation using QUEST2.

Interpreting the results of QUEST2 can be done on a scale by scale basis. It provides a more accurate and comprehensive measure of where weaknesses exist, thus leading to a better interpretation of the steps that need to be taken to ameliorate them.

Our suggestion for the relationship between and administration of QUEST3 and QUEST2 are as follows:

Use QUEST3 as an initial "scan" of the program through calculation of the total score.

Provide an initial interpretation of the score for QUEST3 that leans towards ensuring quality programs for children and is thus very conservative and biased towards a follow-up with QUEST2.

In other words, if a program scores 5 (average) on all items in QUEST3, their total score would be 70. If we want to ensure quality programs for children, we could say that a score of 70 (50%) or less is "not good enough" on the initial scan and that follow-up should be done by administering QUEST2 to more comprehensively measure the program.

We could even be more conservative and say that if a program does not score 80% or more on QUEST3 (i.e., 100 or higher), then more detailed measurement is required. We would lean towards the latter interpretation. In this case, only exemplary programs would not need to administer QUEST2.

5.5 Norms for QUEST2 and QUEST3

In the initial stages of the program, and if possible, an attempt should be made to collect the results for both QUEST2 and QUEST3 (centrally) to see "where programs are at" on both tools, and to begin to develop norms for both. Knowing where programs are situated now will help decide on an appropriate cut-off for QUEST3, as well as the score programs should "shoot for" on QUEST2 for High Five accreditation.

5.6 Final Statement

We do not view this document as the final assessment of QUEST (QUEST1, QUEST2 and QUEST3). We do view it as a first step in understanding the tools' psychometric characteristics, and as process that has raised their measurement credibility. It is also a stepping stone for even more enhancements in the near future.

In fact, we would recommend that the internal consistency reliability, interrater reliability, and construct validity of QUEST continue to be investigated and reviewed on a regular basis as the program unfolds.

Appendix A
Observing the Child's Experience - Comprehensive
(Under Separate Cover)

**Appendix B
Review of Best Practices
(Under Separate Cover)**

Appendix C
Average Ratings of QUEST1 Items by Recreation Providers

Item	Average Rating
A1.1 The hiring criteria of all facility staff/individuals who come into contact with children includes their ability to interact positively with children.	4.38
A1.2 Screening practices are in place for all individuals who come into contact with children.	4.75
A1.3 Police record checks are conducted for all leaders coming into contact with children.	4.00
A1.4 Hiring criteria for program leaders include: interest in children understanding of children and their development a caring attitude toward children experience working with children	4.75
A1.5 References are checked for all program leaders as part of the selection process.	4.63
A1.6 Leaders' screening criteria includes ability to give leadership to children and promote healthy development.	4.50
A2.1 Permitted practices for how children should be treated are clearly documented.	4.38
A2.2 Clear guidelines are in place and communicated to parents regarding how leaders must deal with issues of racism, bullying, sexual harassment and disrespectful behaviour	3.75
A2.3 Prohibited practices including corporal punishment, sexual abuse, deprivation, lack of supervision, humiliation, threats, intimidation and sarcasm are clearly articulated.	4.75
A.2.4 Clear practices are in place and articulated regarding the management and dissemination of confidential information.	4.63
A.3.1 There is collaboration/ cooperation with other facility leaders/ individuals to ensure the best experience for the child.	4.29
A.3.2 There is collaboration/ cooperation with the community and other service providers with the goal of maximizing the number of opportunities and the number of children participating while ensuring continuity of care.	4.00
A.4.1 Leaders are given clear guidelines regarding their responsibilities for healthy child development.	4.50
A.4.2 All facility staff/individuals who come into contact with children receive training on their responsibility to children.	4.00
A.4.3 Job descriptions are available for all positions, clearly articulating leaders' responsibilities with regard to the development of children.	4.63
A.4.5 Leaders' training minimally includes: child development child guidance communication with parents policies and procedures safety and emergency procedures	4.86

Item	Average Rating
B.1.1 Crisis Management procedures define roles and responsibilities including: Containing the crisis Emergency medical procedures Evacuation procedures/drills for all settings and facilities used, both indoor and outdoor Access to phone numbers	4.88
B.1.2 Crisis Management procedures define roles and responsibilities including: Communication processes, both internal and external to parents, media, helping agencies etc.	4.50
B.1.3 There are procedures for dealing with, documenting and reporting emergency situations and both life-threatening and non-life-threatening crises.	4.75
B.1.4 First aid kits are adequately stocked and available at all sites.	4.88
B.1.5 Procedures are in place for dealing with medical emergencies.	5.00
B.2.1 Training on detection, reporting procedures and dealing with disclosures of abuse is conducted as part of leader and volunteer training and orientation procedures.	4.75
B.2.2 Procedures are in place for dealing with and reporting leaders who are suspected of child abuse.	4.88
B.2.3 Legal requirements for dealing with and reporting child abuse are followed.	4.75
B.3.1 Regulations and procedures for dealing with sick children are in place.	4.50
B.3.2 Procedures for administration of prescription and non-prescription medication are in place.	4.75
B.3.3 Health conditions and illnesses under which children may or may not participate are clearly articulated.	4.38
B.4.1 Processes are in place to ensure safe and secure greetings and departures and parents are informed of these procedures.	4.75
B.4.2 Children are supervised appropriately by leaders at all times, including escorting or employing buddy system when children leave the group.	4.88
B.5.1 Attendance is recorded daily.	4.50
B.5.2 The following information is obtained and kept on file for each child. Name, address and phone number Parents' names, addresses and phone numbers Ontario Health Card number Health conditions	4.88
B.6.1 All policies and procedures are communicated and/or available to parents, with particular emphasis on: Child Guidance/Safety	4.13
B.6.2 Procedures are in place for the release of children to adults other than parents.	4.88
B.6.3 A safe arrival program involving a phone call to parents or a parental waiver is in place.	4.50

Item	Average Rating
B.7.1 There is a conscious decision to consider the age of the leaders/ volunteers as part of the selection/ placement process to ensure that the supervision needs of the children and program are met.	4.38
B.7.2 The extent of volunteer involvement is based upon program type and level of required supervision	4.25
B.7.3 Leaders are evaluated at least once per session.	4.38
B.8.1 Sanitary procedures are clearly articulated for all elements of the program. (eg. washing toys and equipment)	4.13
B.9.1 Procedures for safe preparation and serving of food are in place.	4.38
B.9.2 Food served is in accordance with the Canada Food Guide or in consultation with a nutritionist.	3.88
B.9.3 The menu reflects the cultural makeup of the children in attendance.	3.63
C.1.1 Leaders share information with parents regarding their child's experience at least once per session through informal chats, scheduled discussion, written communication etc.	3.75
C.1.2 Written materials on the program are available and written appropriately for both parents and children.	3.88
C.1.3 Parents are actively encouraged to become involved in the program in a manner suitable to the program type. Eg. visiting, assisting, providing feedback etc.	3.87
C.2.1 Parents have the opportunity to give feedback in writing on their child's experience and their level of satisfaction.	4.00
C.2.2 Children have regular opportunities to give feedback on and help plan their experience.	4.38
C.2.3 Programs are formally evaluated at least once per session.	4.50
C.2.4 The evaluation includes assessment based on the healthy development of children plus child and parent feedback.	4.13
C.3.1 Programs are planned according to the Guiding Principles for Healthy Child Development in Recreation and Sport.	3.88
C.3.2 There is a statement of philosophy and goals for children based upon the Objectives for Healthy Child Development in Recreation Settings.	3.86
C.3.3 The planning process for programs includes consultation with other facility leaders/individuals serving these children to ensure the children's best interests are accounted for.	4.00
C.3.4 Procedures are in place to ensure that written program plans are reviewed by a supervisor prior to the start of the program.	4.25
C.4.1 Transitions between one activity to another and one location to another are planned as an integral part of the program and experience.	4.00
D.1.1 All locations comply with applicable provincial and/or municipal health and safety standards. (Eg. Health and Safety code)	4.75

Item	Average Rating
D.1.2 Regular equipment and facility safety checks are made by leaders.	5.00
D.1.3 There is access to phones at all sites	5.00
D.1.4 All leaders complete a pre-program site visit/orientation prior to the commencement of the program.	4.75