REPORT ON THE DEVELOPMENT OF THE ELOM LEARNING PROGRAMME QUALITY ASSESSMENT TOOL

Compiled by Linda Biersteker, Colin Tredoux, Frankie Mattes and Andrew Dawes

28 March 2022

RATIONALE

Early learning programmes can facilitate child learning and prepare children for successful transition to school. However, the quality of these programmes is a significant factor in determining whether and by how much, children will benefit from their early learning experiences. Extensive evidence indicates that programmes of high quality are most effective[1].

A number of early learning programme quality assessment instruments are available and were considered when designing the LPQA Tool. While appropriate for research purposes, many instruments such as the ECERS-R, have some limitations for low resource contexts and where assessments have to be done at scale with limited expertise. First, they have not been developed with South African policy and standards in mind and do not align to the regulatory framework. Second, they require extensive training, and are preferably administered by observers well qualified in early education. Third, they may require extended observations of the classroom environment (several hours). The qualifications of assessors and length of observations makes them costly to implement at scale.

The need for a short, easy to administer, measure of programme quality aligned with the South African curriculum framework, was the impetus for development of the ELOM LPQA tool. This short, generic ELOM LPQA tool is for use by organisations wishing to determine the quality of an Early Learning Programme. The instrument is designed to rate the quality of ELPs in five domains: programme environment, classroom curriculum, learner assessments, relationships and interactions, and teaching strategies.

POTENTIAL USES

- For studies of the quality of pre-Grade R programmes
- To help ECD Resource and Training organisations to interpret factors contributing to children's ELOM domain scores and indicate areas of the learning programme which may need strengthening.
- To use as an element of a quality rating and improvement system and to track improvement over time.
- To provide a reliable learning programme quality score for local, provincial or national samples which can be used to track improvements over time and reporting on the quality component of SDG 4.2.

The tool is for use with group learning programmes (ECD centres and playgroups) delivered directly to children aged 3 -5 years. It focuses **only** on provision and implementation of the learning programme/curriculum activities.

---

[1] Britto, P., Yoshikawa, H. & Boller, K. (2010). Quality of early childhood development programs in global contexts. Rationale for investment, conceptual framework and implications for equity. Social Policy Report, 25 (2); Burchinal,M., Zaslow,M. & Tarullo, L. (2016). (Editors).Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes. Monograph: Society for Research in Child Development; Rao, N., Sun, J., Wong, J.M.S., Weekes, B., Ip, P., Shaeffer, S., Young, M., Bray, M., Chen, E. & Lee,D. (2014). Early childhood development and cognitive development in developing countries: a rigorous literature review. London: Department for International Development; Sabol, T. J., Hong, S. S., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning?. Science, 341(6148), 845-846.

This is a short instrument and does not provide detail about all the equipment and activities offered, but focuses more generally on the variety of activities and material and teaching strategies. Mathematics and Language as key predictors for Grade R are the exception.

SELECTION OF ITEMS FOR LPQA

Selection of items was informed by a review of the literature and drew on the key classroom focus areas of established observational measures which will support construct validity.

While definitions of high quality ECCE vary somewhat according to context, and include many factors beyond the classroom, there is general agreement that the learning environment set up and resourcing, curriculum, approaches to extending learning, teacher child interactions and relationships are critical contributors to programme quality and child learning outcomes.

Following the literature, well known measures of quality, local experience as well as South Africa's National Curriculum Framework with particular focus on the Towards Grade R phase, and National Early Learning and Development Standards, a short list of items was drafted to represent the following areas:

1. Presence in the learning environment of a variety of activities supported by materials and books (Aboud, 2006; Montie, Xiang & Schweinhart, 2006; Trawick-Smith et al, 2015; UNESCO, 2017).
2. Implementation of a holistic age-appropriate curriculum. To promote school readiness there should be a targeted focus on specific school readiness skills (early mathematics and literacy) with clear learning goals, rather than a general whole child curriculum which includes these skills. Effective learning activities should be cumulative and sequenced to align with children's developmental stages (Center on the Developing Child, 2016; Phillips et al., 2017; UNESCO, 2017).
3. Rich language and literacy experiences are the basis for learning and later reading (Lonigan et al., 2000; Opel et al 2009; Storch & Whitehurst, 2002).
4. Opportunities for child-initiated activities individual and with peers as well as adult-led individual and small and whole group activities (Jenkins & Duncan, 2017; Phillips et al, 2017; Montie et al. 2006; Sylva et al., 2007).
5. Social and emotionally supportive relationships (including fostering of independence and self-regulation (Durlak et al, 2011; Shala, 2013; Wolf et al,2018).
6. Sensitive, mediated caregiver/child interaction targeted to the developmental levels and needs of individual children (AKF, 2010; Sylva et al., 2007; UNESCO, 2017), and designed to address areas that need strengthening. This requires careful planning and assessment of individual abilities (Grisham-Brown, Hallam & Brookshire, 2006).
7. Play promotes learning and development. A continuum of different types of play provide for this, from that which is freely chosen by children, through adult-guided play (in which adults scaffold child-led play), to adult-structured activities where the teacher designs, sets rules and scaffolds play with a particular learning objective (Edwards and Cutter- Mackenzie, 2013; Jensen, Pyle, Zosh, Ebrahim et al., 2019; Pyle and Danniels, 2017; Zosh et al, 2018). Highly teacher-controlled, direct instruction methods, such as large group worksheet-based academic activities should be avoided as they have been linked with stress and reduced motivation in preschool children (Elkind, 1986; Stipek et al., 1995).

Items selected had to:

- Be easily observable (direct observation or documentary records) and with differentiated levels for scoring. Be strongly associated with overall quality scores and child outcomes related to readiness to benefit from Grade R.
- Align with the SA National Curriculum Framework curriculum aims - towards Grade R competencies and the ECD Programme sections of the Children's Act ECD Norms and Standards.

Items selected are frequently used in classroom quality observations. Table 1 in Appendix 1 summarises international and local measures which include items relating to the subscales of the LPQA.

PSYCHOMETRY

In March 2020, a draft LPQA was piloted in 130 early learning programmes before fieldwork was halted due to the Covid-19 lockdown of early learning programmes. Data from this pilot was used to conduct Exploratory Factor Analysis to determine whether the items showed internal validity and reliably measured the same construct of "quality". As a result, one question from the Curriculum sub-scale was dropped as it displayed low levels of internal consistency with the rest of the items in the tool.

Between September and November 2021, the revised LPQA was administered in over 500 randomly selected early learning programmes across all nine provinces in South Africa as part of the Thrive by Five Index. For each of these ELPs child outcomes data was also collected (±4 children per site).

In preparation for data collection, nineteen observers were trained over three days during which they were introduced to the items and familiarised with examples. Videos used to help establish appropriate ratings for the different levels of each item. Trainee observers then observed two different early learning classes in small groups together with an expert observer and each independently completed the LPQA. Assessor ratings were compared and inconsistencies discussed. There were few deviations for subscale means between the trainee observers and experts.

Confirmatory Factor Analysis (CFA) was conducted on 477 records, after data cleaning (See Appendix 2 for details of psychometric procedures). Five factors (subscales) with good item fit on several commonly used indices, and on which items had satisfactory coefficients, constitute the subscales and items of the ELOM LPQA.

1. The Learning Environment (5 items)
2. Assessment for Learning and Teaching (2 items)
3. Relationships and Interactions (4 items)
4. Curriculum (5 items)
5. Teaching Strategies (5 items)

Gross Motor materials and activities is an additional stand-alone item that did not load on the Learning Environment factor but is included as it is important for monitoring equipment for large muscle development. The score on this item does not contribute to any subscale or the LPQA Total score.

The overall conclusion is that four of the five scales correlate with criterion variables as we expected them to. The only scale that did not do so is that measuring relationships, and it might well be appropriate that it did not correlate with the criterion variables, given the nature of the items in the scale.

The recommendation was thus to keep the five subscales as originally devised, with some minor modifications which have since been made to the tool.

Each item in the final set is:

- Easily observable (direct observation or documentary records) and with differentiated levels for scoring
- Strongly associated with overall quality scores and child outcomes related to readiness to benefit from Grade R
- Aligned with the SA National Curriculum Framework – towards Grade R phase competencies and the Programme section of the Children's Act ECD Norms and Standards.

SCORING THE LPQA

Each item is scored on a scale of 0 (inadequate), 1 (basic) or 2 (good) to provide some range of scores, but also to take into account that more nuanced and extended scales require experienced and well-qualified ECD assessors.

The discipline item of the *Relationships and Interactions* subscale includes a 'not observed' category as that item may not be seen during the limited period of observation (in this case the scale score must be pro-rated). Explanatory notes have been added to items.

While a more extended scale (e.g. 7 points) would have been better practice for this sort of tool (as, for example, in the ECERS-3), we chose a 3 point scale as in our experience in South Africa, finer gradations are often challenging for observers who are not experienced ECD trainers, and in particularly, for field staff used in large surveys. This renders their scores less reliable.

- **Deriving Subscale Total Scores**: Scores on all items in each subscale are summed to derive a Total Subscale Score. Five subscale total scores can thus be derived. These are the most important scores in the LPQA as they show the strengths and weaknesses of the programme and are of assistance when providing guidance for improvement.
- **Deriving the LPQA Total Score**: Subscale Total Scores are summed to derive this score. This is common practice in similar tools such as the ECERS. It must be remembered that the subscales comprise different numbers of items and that variation in performance in each subscale will affect the Total. That is why Subscale Total Scores must be examined to contribute to understanding the LPQA Total Score.

ADMINISTRATION OF THE LPQA

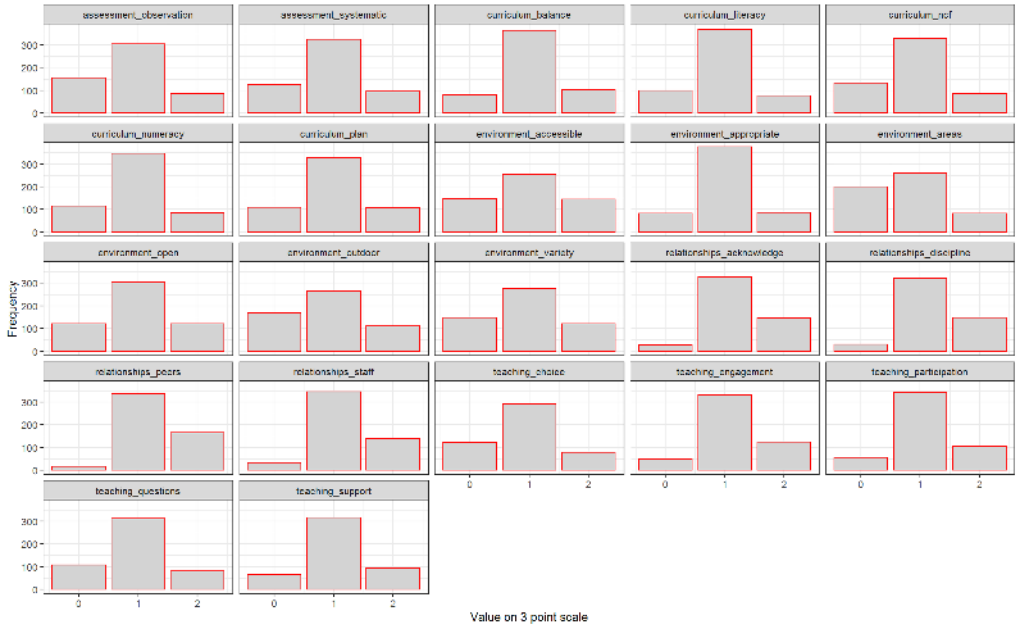
Training is essential for all users, but particularly for those who are not experienced ECD trainers, to reduce subjective interpretation and ensure sound inter-rater reliability. A draft training manual has been compiled (see attached document).

The LPQA tool is used at classroom level, and if there is more than one classroom, each will require a separate rating. Observation time per class, including reference to planning or child records, should be for a minimum of **two hours** at a time of day that free choice or small group activities (indoor playtime) **and** at least one large group activity (morning ring, story or music time) can be seen. Documentary records such as planning sheets, child records and daily schedules should be included in the assessment.

Items need not be rated in the order of the user manual but rather as they are seen. Materials and layout can be assessed during routines such as meals and toilet or prior to the start of the daily programme. Relationships and teaching strategies will be seen throughout the observation period and ratings should be finalised at the end of the observation period.

INITIAL FINDINGS ON THE PROGRAMME QUALITY COMPONENT OF THE THRIVE BY FIVE STUDY

Figure 1 shows the distribution of individual items and what can be seen is the modal (highest frequency) rating for all items was at the basic of quality but there was some variation for all items (Items on the scale are scored 0 (inadequate), basic (1) or good (2)).



The table below provides the means for each of the Subscales and a description of each -

Subscale	Mean	Description
Learning Environment	0,95	**Learning Environment** provides a rating of how the playroom has been divided into different activity areas and the availability of a variety of materials that are age appropriate, accessible to the children and include open ended materials which promote the use of imagination and problem solving. Materials are important because they support the curriculum and play based learning opportunities. The average score was 0,95 out of 2. The lowest average for the availability of different activity areas which underpins free choice and a play based approach.
Relationships and Interactions	1,24	**Relationships and Interactions** refers to the classroom climate. To be conducive for learning, a warm responsive practitioner who acknowledges efforts, encourages positive social relations between children and uses positive discipline to maintain classroom control and teach prosocial behaviour is desirable. This subscale had the highest mean at 1,24 out of 2. This finding is similar to that on the ECERS -R Interactions Subscale which tends to find that quality of classroom climate is supportive and warm and child centred. [2,3]
Curriculum	0,97	**Curriculum** covers use of the National Curriculum Framework to guide the programme, a regular well balanced daily schedule, evidence of planning and provision of numeracy and mathematics activities and language and literacy promoting activities. The subscale average was 0,97 out of 2. The lowest rating was for use of the NCF to guide

[2] Biersteker, L., Dawes, A., Hendricks, L., & Tredoux, C. (2016). Center-based early childhood care and education program quality: A South African study. Early Childhood Research Quarterly, 36, 334–344. https://doi.org/10.1016/j.ecresq.2016.01.004

[3] Mwaura, P. A. M., Sylva, K., & Malmberg, L. (2008). Evaluating the Madrasa preschool programme in East Africa: A quasi-experimental study. International Journal of Early Years Education, 16(3), 237–255. https://doi.org/10.1080/09669760802357121

		the programme. Understanding and use of the NCF has been noted as a gap in other studies.[4]
Teaching Strategies	1,03	**Teaching strategies** refers to the way that practitioners scaffold and extend learning through questions and support children, as well as the extent to which they enable free choice and autonomy. These strategies have been shown in the literature to be most strongly associated with learning outcomes. The average for this scale was 1,03 out of 2 which indicates that practitioners are implementing these at a basic level. Choice of activities was the lowest item which is likely to reflect covid protocols which restricted choice to maintain social distancing.
Assessment	0,91	**Assessment for learning and teaching** contains only two items, observing individual children and noting their strengths and weaknesses to inform interventions and systematic regular progress assessments and reports. Being responsive to individual learning needs, early identification of difficulties are a critical aspect of quality. At 0,91 this was the lowest subscale average. The Department of Basic Education with UNICEF support has recognized the need for more training and tools in this regard and have drafted an assessment package in support of the NCF which will be available shortly.

Preliminary explorations of the relation of different subscales to outcomes include:

- Environment correlates significantly with socio-economic measures (subsidy, quintile, fees), and with learning outcomes as measured by the total ELOM 4&5 score.
- Curriculum correlates with socio-economic variables
- Teaching strategies correlates with learning outcomes as measured by the total ELOM 4&5 score
- Assessment correlates with both socio-economic variables and the total ELOM 4&5 score.
- The only scale that did not correlate with these variables is Relationships which may well be explained by the fact that it is an underpinning factor for social and emotional development which was not measured in this study.

NEXT STEPS

- **Further analysis** - Further analysis is planned to review the relationship between other characteristics of the ELP (e.g. practitioner qualifications, levels of experience etc.) and the LPQA scores, as well as between LPQA scores and child outcomes.
- **Promote access to the LPQA tool** - the following activities are underway to facilitate the use of the LPQA tool and the data generated -
    - completion of the draft training manual (attached)
    - final changes to the LPQA user interface on Survey CTO
    - backend systems to enable automated data cleaning and reporting (enables affordable use at scale)
    - production of communications assets to raise awareness of the tool and its potential value-add
- **Facilitate use of data insights to drive change** - As with all ELOM tools, the intention is to provide users with feedback in ways that enable clear and evidence-informed programme enhancements. Feedback typically includes the opportunity to engage with an expert to brainstorm quality enhancement strategies, and the provision of appropriate resources matched to the strengths and weaknesses of each programme. Next steps will include testing of various feedback mechanisms.
- **Open access data** - to encourage further analyses, we aim where possible to add anonymised datasets to an open access repository. This enables and encourages researchers from diverse backgrounds to undertake secondary analyses which often yield new and valuable insights.

THE LPQA SUBSCALES AND ITEMS

The updated subscales and items, and the overall scoresheet, are included in the tables that follow.

[4] Biersteker, L. (2021) Practitioners' perceptions and understanding of the approaches underpinning curriculum and pedagogy in an early childhood classroom. Pretoria: Umalusi

SUBSCALES AND ITEMS

LEARNING ENVIRONMENT

1. Room arrangement	Inadequate 1	Basic 2	Good 3	Terms and examples
During playtimes the room is divided into learning areas/ interest centres	There are no or fewer than three organised learning areas.	Three learning areas set out for children to play in (alone or with other children).	There are four or more learning areas arranged so children can use them, quiet and active areas are separate.	**Learning areas/interest centres** refer to spaces organised with equipment for different kinds of play. These can include make-believe play, big blocks, sensory play (sand, water etc) an art area (painting, drawing, modelling, cutting and pasting, box construction etc) book area, nature and science table, educational toys and games (puzzles, small construction, sorting games, counting, threading etc) More than one make-believe activity counts as **one** area e.g. home and shop. Quiet areas include books, educational toys and games, fine motor and art and active include blocks and make believe.

2. Indoor Materials	Inadequate 1	Basic 2	Good 3	Terms and examples
There are enough and varied materials for play and learning indoors	Insufficient materials in any or all areas e.g. children have to wait for a toy or resource and have nothing to work with while waiting.	Some materials for all learning areas e.g. children don't have to wait but there is limited choice – only one activity or only one choice.	Enough and variety of materials in all learning areas - more than two activities in each area and enough for all children to be occupied.	**Enough** means that there is sufficient equipment for all children to have a choice of activities at all times(allowing for rotation and turn taking). **Choice** refers to a variety of things to do within an area e.g. several books, drawing and painting, lego and table blocks etc.

3. Developmentally appropriate materials	Inadequate 1	Basic 2	Good 3	Terms and examples
Materials for play and learning are developmentally appropriate for 4 – 5 year olds	Many materials are not appropriate for the developmental level and interests of the child.	Most materials match the developmental level and interests of children 4 – 5 years (e.g. children are engaged, some easier and some more challenging materials).	Materials provide for a range of developmental levels and interests including adaptive equipment for children with special needs (e.g. many interesting things to do, suitable for a range of abilities).	Developmentally appropriate means that the materials accommodate the learning needs of this age group and include some easier and more challenging materials to meet individual needs. Adaptive for special needs may include large format books for visual impairment, easy to handle puzzle pieces and communication boards.

4. Accessible, materials	Inadequate 1	Basic 2	Good 3	Terms and examples
Materials are accessible for children.	Few or no materials are accessible to children.	Some materials are easily accessible for children.	All materials are laid out so that children can easily access them.	**Accessible** means that children can easily select materials themselves – they are not packed away.

5. Open ended materials	Inadequate 1	Basic 2	Good 3	Terms and examples
There are **open ended materials**	No open ended materials.	At least one example of open ended materials in each of two learning areas.	At least two examples of open ended materials in each of three learning areas.	Open ended materials are those that can be used in many different ways e.g. blocks, cloths, free art materials, boxes, sand and water, natural materials like stones, seed pods, sticks. These allow children to use their imagination, problem solving and creativity and create their own play experiences. Worksheets and pre- prepared craft projects where all children do the same thing are not open ended.

6. Child observation	Inadequate 1	Basic 2	Good 3	Terms and examples
Children are observed to inform planning and support needs	There is no observation record or the observation book is not in regular use.	There is a regular observation record/book with some written indication of remedial activities.	There are many and varied observations of children's progress and evidence of a range of activities to remediate difficulties/facilitate holistic development.	Ask the teacher/practitioner to show you any record of observations of learner performance (e.g. observation book).

7. Recording progress	Inadequate 1	Basic 2	Good 3	Definitions
Each child's progress is regularly and systematically monitored in an informal and play-based way.	There is no systematic child assessment.	Assessments are done at least twice a year, using a standard format, and filed for reference.	Assessments using a standard format are updated regularly, children with difficulties assessed more often	Ask to see copies of any assessment forms, child progress or development records, child portfolios. If there are records, ask how often progress is assessed and recorded.

RELATIONSHIPS AND INTERACTIONS

8. Child-child interaction	Inadequate 1	Basic 2	Good 3	Terms and examples
Practitioner promotes and encourages positive interaction among children	Children are discouraged from interacting (e.g. not allowed to choose who to play with, focus on individual work) and there is little or no guidance for positive peer interaction.	Peer interaction encouraged (e.g. free choice of who to play with), negative interactions stopped.	Staff help children develop good social behaviours, provide activities that encourage children to work together, support children who find it hard to join in.	Examples of promoting positive child/child interaction include encourage playing and working cooperatively, helping children who find it hard to join in to join a group).

9. Staff child interaction	Inadequate 1	Basic 2	Good 3	Terms and examples
Staff child interaction warm, interested individualised and respectful, sensitive to how children are feeling	Staff are not responsive to or involved with children (ignore or just give instructions) little individual attention.	Friendly atmosphere, some positive interactions and response to individual children, consistent response to children's needs (observed at least twice).	Frequent positive interaction, warm contact, relaxed and pleasant atmosphere Sensitive to non-verbal cues and respond appropriately, respectful and guide positively, supportive and comforting (observed	This means that staff notice what individual children are doing and feeling and act accordingly. For example, they respond promptly to children who are upset and notice when children need assistance.

			more than twice during observation.	

10. Child efforts acknowledged	Inadequate 1	Basic 2	Good 3	Terms and examples
Children's efforts and ideas acknowledged.	Staff do not use encouragement to acknowledge children's efforts or ideas.	Staff sometimes use encouragement (at least two instances observed with two different children).	Staff regularly use encouragement to acknowledge individual children's efforts and ideas.	Acknowledgement includes practitioner behaviours such as repeating child's ideas, comments on what children are doing, putting them in control of evaluating their own work and efforts. To get a rating of 2 there needs to be more engagement with the child than a statement such as well done, high five or good job

11. Discipline	Inadequate 1	Basic 2	Good 3	Not seen	Terms and examples
Positive discipline	No positive discipline observed or expectations may not be age appropriate.	No harsh methods used, control maintained, expectations age appropriate.	Effective use of non-punitive methods, staff apply rules consistently, children supported to solve conflicts for themselves, express how they are feeling.	Not seen is the rating if there are no instances requiring behaviour management during the observation period.	There should be no physical punishment, yelling, shaming, withholding food or time out/naughty corner longer than 4 – 5 minutes. Positive discipline involves, setting clear expectations, praising good behaviour. When a child misbehaves staff remind them of rules, explain and redirect unsuitable behaviour, discusses with children etc.

CURRICULUM (PLANNED EXPERIENCES AND ROUTINES)

12. Use of National Curriculum Framework (NCF)	Inadequate 1	Basic 2	Good 3	Terms and examples
Learning programme is guided by the Early Learning and Development Areas (ELDAs)and aim and phase specific developmental guidelines	No evidence that programme supports NCF aims.	Learning programme focuses on some of the ELDAs and is mostly developmentally appropriate (4 – 5 years).	Learning programme used covers all the ELDAs and activities are developmentally appropriate catering for different children's individual needs.	ELDA aims support: Well- being (health and motor development), identity and belonging (social and emotional), communication, exploring mathematics, knowledge of the world, creativity. For this age group the Phase is Towards Grade R. *Check planning book for the day and the activities on offer.*

13. Programme planning	Inadequate 1	Basic 2	Good 3	Terms and examples
Practitioner organises activities according to an integrated weekly and daily plan	There is no evidence of planning used to organise learning activities (that a specific plan is being followed for the day though there may be a regular schedule).	Planning books and the playroom reflect a planned and integrated approach across different learning areas and parts of the daily programme.	Plans are applied and there is evidence of taking into account children's interests and developmental appropriateness in planned activities that are implemented.	An integrated plan will have a focus/ theme/topic shown across learning areas and times of the day). E.g. If family is the topic, it will be discussed as part of morning ring, children may draw or paint family members, there may be songs and a story about different families. *Ask to see planning book/file/notes*

14. Daily programme	Inadequate 1	Basic 2	Good 3	Terms and examples
Programme/daily schedule includes a balance of free play, small group and whole group	Either free play or whole group activities (where children all do the same thing) predominate.	Both free play and whole group activities are provided each day and there are some small group times.	There is a substantial free play indoors and outdoors, at least one story, music and ring time daily and regular small group teaching opportunities with all children are exposed at least once a week.	D: Review daily schedule – this is usually displayed on the wall. Free play where children have choices can be indoors and outdoors. Small groups are organized times for a few children with the practitioner. Whole groups include morning ring, story, music and group games.

15. Numeracy and Mathematics	Inadequate 1	Basic 2	Good 3	Terms and examples
Programme includes numeracy and mathematics activities	Few or no appropriate maths activities, staff do not show children how to use them or participate and practitioner does not use math words when talking to children in daily events, group times.	Some practitioner initiated and directed maths activities including number songs and rhymes linked to an intended purpose and some use of maths words in daily events and routines.	Frequent number songs, rhymes, games. Children encouraged to count objects, name shapes, sort and match. Maths learning also integrated as part of daily routine, how many cups for dolls, plates for children, number wearing red.	Numeracy and mathematics activities and materials include numbers, shapes, measurement, grouping and sorting. There should be a mix of activities especially with concrete materials and reference to maths concepts in all parts of the programme including story, music, serving snack, etc.

16. Language and Literacy	Inadequate 1	Basic 2	Good 3	Terms and examples
Programme includes language and literacy activities	Few appropriate language activities e.g. Practitioner rarely reads to children, few accessible books, little encouragement for children to talk to practitioner, no labelled pictures or print other than books, limited access to writing implements.	Some of the following language/literacy activities – daily story, some appropriate books and reading to children, access to writing implements and paper, some practitioner and child conversations and some labelled pictures and printed words displayed, especially children's names.	Daily story with active child participation and discussion (e.g. children asked to recall parts of story); access to many appropriate books and are read to regularly. Practitioner records children's sayings, labels items in drawings, displays their emergent writing. Children encouraged to answer questions in extended way, individual adult/ child conversations and activities planned to stimulate conversation. Many labelled pictures and materials on view.	Language activities include daily story time, books and reading to children, talking and listening, encouraging conversation encouraged, a print rich environment , opportunities for drawing/writing. Check the environment for labelled objects and children's work, availability of writing materials and children's portfolios.

TEACHING STRATEGIES

17. Free choice	Inadequate 1	Basic 2	Good 3	Terms and examples
During free choice times children have a high level of choice about what to play and what materials to use	Staff direct how children use materials and carry out activities (e.g. all make same things, respond with same words and actions).	Children make at least two choices independent of practitioner direction about where and how to carry out activities, but some materials choices and activities are practitioner directed.	Children make three or more choices during playtime (independent of practitioner direction) about where and how to use materials and carry out activities.	Practitioners do not give instructions to children concerning their choice of activities or playmates.

18. Staff child engagement	Inadequate 1	Basic 2	Good 3	Terms and examples
Staff move around and engage with children during playtime and use a range of techniques to support and extend children's learning	Staff do not engage to support and extend children's learning. Interaction is largely supervisory.	Staff engage with one or two children to extend their learning using one or two different techniques (3 instances observed).	Staff regularly engage to extend children's learning using a variety of techniques (more than 3 instances observed).	Techniques to extend learning may include conversation in which children talk and practitioners listen, adding information, questioning, modelling how to do something, joining in play, allowing children to try things out, simplifying a task or suggesting a more advanced activity.

19. Group times	Inadequate 0	Basic 1	Good 2	Terms and examples
Group times which are practitioner directed allow for child participation.	In large group activities staff do not ask children to offer ideas or participate	Children sometimes contribute or participate at own	Children contribute their own ideas and participate at own developmental levels	Group times are times when all children are engaged in an activity organised by the practitioner – ring time, music, story etc.

	according to their developmental levels.	developmental levels at large group time.	throughout large group times.	

20. Questioning to extend learning	Inadequate 0	Basic 1	Good 2	Terms and examples
Staff ask open ended questions to extend children's thinking	No open-ended questions to extend children's thinking.	Some questions to encourage children to reflect on an activity or idea (2 instances observed).	Many questions to encourage children to reflect on actions and ideas in multiple ways (3 instances observed).	Open ended questions are those that go beyond a question to which there is only one answer. e.g. What colour is this? They encourage further thinking e.g. 'What do you think ….', 'Do you agree with …?', 'Why?', 'Would you have done it differently? How?

21. Emotional development [5]	Inadequate 1	Basic 2	Good 3	Terms and examples
Staff support children's emotional development: opportunities for autonomy	Staff do not support children to do things for themselves or encourage children to take initiative.	Practitioner sometimes supports children to do things for themselves and take initiative (2 instances observed).	Practitioner regularly supports children to do things for themselves and take initiative (3 or more instances observed).	Examples of helping children to be independent would be encouraging them to do dress themselves, pour juice, fetch paper, help tidy up, and taking initiative (trying things in different ways, suggest a game etc).

[5] Staff support for emotional development had a coefficient of .49 and the acceptable coefficient is .50 as noted in the Psychometric Appendix. However it is likely that ratings were affected by Covid social distancing protocols and many missing values during the time it was piloted. It is an important NCF goal and emotional functioning is significantly associated with child learning outcomes, so it has been retained.

The following item does not form part of any of the subscales and LPQA total score but has been included as a stand-alone item to ensure that provision is made for Gross Motor Development.

Gross Motor materials and activities	Inadequate 0	Basic 1	Good 2	Terms and examples
Gross Motor materials and equipment to encourage development of different motor skills	Few or no gross motor equipment/opportunities.	Some gross motor equipment to promote different kinds of movement.	A variety of small and large equipment to promote different kinds of movement.	Motor skills include running, balancing, swinging, hopping, skipping, climbing, throwing and catching, managing wheel toys etc). Equipment to encourage Gross Motor Development includes e.g. any of the following: small equipment such as skittles, beanbags, ropes, balls, tyres, large fixed equipment such as climbing frames or swings, commando nets. This will mostly be outside but may be in a space where children can move freely.

SCORE SHEET

Date:		Observer:	
Time observation started:		Time observation ended:	
ECD Programme Name:		Class/Group:	
Age range of children in the class (in months)		Number of children enrolled in the class:	
Number of children present during observation:		Number of teaching assistants:	

LEARNING ENVIRONMENT	1 Inadequate	2 Basic	3 Good	Notes
1.Room arrangement				
2.Indoor materials				
3.Developmentally appropriate materials				
4.Accessible, safe materials				
5.Open-ended materials				

Learning Environment Score: Sum of items ÷ number of items _____

ASSESSMENT FOR LEARNING AND TEACHING	1 Inadequate	2 Basic	3 Good	Notes
6.Child observation				
7. Progress records				

Assessment for Learning and Teaching Score: Sum of items ÷ number of items _____

RELATIONSHIPS & INTERACTIONS	1 Inadequate	2 Basic	3 Good	Notes	
8.Child-child interaction					
9. Staff-child interaction					
10. Child efforts acknowledged					
11. Positive discipline					Not Observed

Relationships and Interactions Score: Sum of items ÷ number of items _____

CURRICULUM	1 Inadequate	2 Basic	3 Good	Notes
12.Use of NCF				
13. Programme Planning				
14. Daily schedule/programme				
15. Language and literacy				
16.Numeracy and mathematics				

Curriculum: Sum of items ÷ number of items_____

TEACHING STRATEGIES	1 Inadequate	2 Basic	3 Good	Notes
17.Free play				
18.Staff child engagement				
19.Group times				
20.Questions to extend learning				
21.Emotional development				

Teaching strategies Score: Sum of items ÷ number of items _____

ELOM LPQA TOTAL SCORE: Sum Total scores of each of the 5 subscales: ____

REFERENCES

1. Aboud, F. E. (2006) Evaluation of an early childhood preschool programme in rural Bangladesh. *Early Childhood Research Quarterly 2,* 46 – 60.
2. Aga Khan Foundation (AKF). (2010) *Improving learning achievement in early primary in low-income countries: A review of the research.* Geneva, Switzerland.
3. Britto, P., Yoshikawa, H. & Boller, K. (2010) Quality of early childhood development programs in global contexts. Rationale for investment, conceptual framework and implications for equity. *Social Policy Report*, *25* (2).
4. Burchinal,M., Zaslow,M. & Tarullo, L. (2016). (Editors).*Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes*. Monograph: Society for Research in Child Development.
5. Center on the Developing Child at Harvard University. (2016). *From best practices to breakthrough impacts: A science-based approach to building a more promising future for young children and families.* Retrieved from www.developingchild.harvard.edu.
6. Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x
7. Elkind, D. (1986). Formal education and early childhood education: An essential difference. *The Phi Delta Kappan* 67, no. 9 (1986):631-636. Cited in Jenkins & Duncan (2017).
8. Edwards, S. & Cutter Mackenzie, A. (2013). Pedagogical play types: what do they suggest for learning about sustainability in early childhood education? *International Journal of Early Childhood, 45*, 327 – 346.
9. Grisham-Brown, J., Hallam, R. & Brookshire, R. (2006) Using authentic assessment to evidence children's progress toward early learning standards*. Early childhood Education Journal*, **34**(1), 45-47.
10. Hamre, B. & Pianta, R.C. (2005) Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development 76* (5) 949-967.
11. Jenkins, J. M. & Duncan, G.R. (2017). Do pre-kindergarten curricula matter? The current state of scientific knowledge on pre-kindergarten effects. In Phillips, D., Lipsey, M., Dodge, et al.. Washington DC: Brookings Institute p 37 – 44
12. Jensen, H., Pyle, A., Zosh, J. M., Ebrahim, H. B.,Scherman, A. Z., Reunamo, J., & Hamre, B. K. (2019).*Play facilitation: the science behind the art of engaging young children (white paper).* Denmark: The LEGO Foundation
13. Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*(5), 596.
14. Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, *21*(3), 313-331.
15. Opel, A., Ameer, S. S., & Aboud, F. E. (2009). The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children. *International Journal of Educational Research*, *48*(1), 12-20.
16. Phillips, D., Lipsey, M., Dodge, K., Haskings, R., Bassok, D. et al. (2017). *Puzzling It Out: The current state of scientific knowledge on pre-kindergarten effects- A consensus statement.* Washington DC: Brookings Institute.
17. Pyle, A. & Danniels, E. (2017). A continuum of play-based learning: The role of the teacher in play-based pedagogy and the fear of hijacking play. *Early Education and Development 28,* 3, 274–289.
18. Rao, N., Sun, J., Wong, J.M.S., Weekes, B., Ip, P., Shaeffer, S., Young, M., Bray, M., Chen, E. & Lee, D. (2014). *Early childhood development and cognitive development in developing countries: a rigorous literature review*. London: Department for International Development.
19. Sabol, T. J., Hong, S.L., Pianta, R.C.& Burchinal, M. R. (2013) Can Rating Pre-K Programs Predict Children's Learning? *Science, 341, issue 6148*, pp 845 – 846,
20. Shala, M. (2013). The impact of preschool social-emotional development on academic success of elementary school students *Psychology (4).*
21. Stipek D., Feiler, R. Daniels, D. & Milburn S. (1995), Effects of different instructional approaches on young children's achievement and motivation. *Child Development 66* (1), 209-223.

22. Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Steven, K., Gilden, R., et al. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education, 15*(1), 49–65.
23. Trawick-Smith, J., Wolff, J., Koschel, M. et al. (2015) Effects of toys on the play quality of preschool children: Influence of gender, ethnicity, and socioeconomic status. *Early Childhood Education Journal 43*: 249.
24. UNESCO (2017). *Overview MELQO (Measuring early learning quality and outcomes).* Paris: UNESCO.
25. Wolf, S., Raza, M., Kim, S., Aber, J. L., Behrman, J., & Seidman, E. (2018). Measuring and predicting process quality in Ghanaian pre-primary classrooms using the Teacher Instructional Practices and Processes System (TIPPS). Early Childhood Research Quarterly, 45, 18–30. https://doi.org/10.1016/j.ecresq.2018.05.003
26. Zosh, J. M.; Hirsh-Pasek, J., Hopkins, E. J., Jensen, H., Liu,C., Neale, C. , Solis, L. & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology, 9*, Article 1124.

APPENDIX 1: ELOM LEARNING PROGRAMME QUALITY ASSESSMENT ITEMS AND SOURCES

The table indicates different Programme Quality Assessments that have related subscales/items.

ELOM LPQA Subscale and Items	Sources
Learning Environment (room arrangement, teaching and learning materials)	ECERS-3[6]Space and Furnishings; Learning Activities Measure of early learning environments(MELE)[7] Play NAEYC Classroom Observation [8] High Scope Preschool Programme Quality Assessment (PQA)[9] Learning Environment DBE National ECD M&E framework Classroom observation tool [10] Classroom and resources
Assessment for Learning and Teaching (observation, progress reports)	ISSA[11] Assessment & Planning NAEYC Classroom Observation
Relationships and Interactions (child-child, adult -child, discipline)	ECERS- 3 Interaction CLASS Pre K[12] Emotional Support, Classroom Organization ISSA Interactions NAEYC Classroom observation High Scope Preschool Programme Quality Assessment (PQA) Adult child interaction Measure of early learning environments(MELE) Interactions Teacher Instructional Practices and Processes System (TIPPS) [13] Emotional support , Classroom organisation. DBE National ECD M&E framework Classroom observation tool: Management of Active Learning
Curriculum (Use of National Curriculum Framework, programme planning, daily schedule/programme, language and literacy, numeracy and mathematics	ECERS- 3:Learning Activities, Language and Literacy, Programme Structure ECERS – E[14] CLASS Pre K Classroom Organization ISSA Assessment & Planning NAEYC Classroom observation Measure of early learning environments(MELE)Pedagogy NAEYC Classroom observation High Scope Preschool Programme Quality Assessment (PQA) Daily Routine

[6] Harms, T., Clifford, R. M., & Cryer, D. (2015). Early Childhood Environment Rating Scale (ECERS), 3rd edition. New York: Teacher's College Press.

[7] https://www.ecdmeasure.org/what-is-melqo/; https://www.ecdmeasure.org/beqi/

[8] National Association for the Education of Young Children. (2018). Early learning accreditation standards and assessment items. Washington DC: NAEYC.

[9] High Scope (n.d.). The preschool programme quality assessment (PQA). Ypsilanti, Michigan: High Scope.

[10] UNICEF & DBE. (2019). National early childhood development monitoring and evaluation framework: South Africa. Pretoria: Khulisa Management Services.

[11] International Step by Step Association (2010) Competent educators of the 21st century: Principles of quality pedagogy. Amsterdam: ISSA.

[12] Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom assessment scoring system: manual pre-K. Education Review// Reseñas Educativas

[13] Seidman, E., Raza, M., Kim, S., & McCoy, J. M. (2014). Teacher instructional practices and processes system–TIPPS: Manual and scoring system. New York: New York University.

[14] Sylva, K., Siraj-Blatchford, I. & Taggart ,B. (2003). Assessing quality in the early years: Early Childhood Environment Rating Scale: Extension (ECERS-E), Four Curricular Subscales. Trentham Books.

	AECI Global Guidelines Curriculum content DBE National ECD M&E framework Classroom observation tool: Management of Active Learning
Teaching strategies (free play and adult directed, scaffolding strategies, support for emotional development).	CLASS Pre K Instructional Support High Scope PQA: Daily Routine, Adult child interaction ISSA Teaching Strategies NAEYC Classroom observation Measure of early learning environments(MELE)Play; Pedagogy Teacher Instructional Practices and Processes System (TIPPS) Facilitating Deeper Learning, Supporting student expression AECI Global Guidelines Pedagogy

APPENDIX 2: TOOL PSYCHOMETRY

Factor structure, internal consistency, and criterion validity of the
ELOM Learning Programme Quality Assessment Tool

CG Tredoux and Frances Mattes, March 2022

In 2020 a preliminary analysis of the LPQA tool was undertaken based on data collected from approximately 130 ELPs. This fieldwork was ultimately curtailed due to Covid and there was insufficient data to finalise the tool. Fieldwork to finalise the psychometrics of this tool was completed as part of the Thrive by Five Index 2021.

As part of the Index fieldwork, a random selection of ELPs were audited to assess programme quality using the draft ELOM Programme Quality Assessment Tool. For each of these ELPs child outcomes data was also collected (±4 children per site).

The data file provided for the analysis, 'thrive_audit_anon.dta' contained data for 571 distinct ELPs, although missing data meant that complete data was only available for 477 of these ELPs.

The report completed with the 2020 data was based on a small sample, and we therefore do not try to repeat or confirm that report, although we do comment on some similar aspects of the factorial validity of the LPQA.

We start by examining some descriptive statistics

***Table 1**: Descriptive statistics for Items in LPQA*

Item	mean	sd	n	n_missing
environment_areas	0.78	0.69	545	26
environment_variety	0.95	0.70	545	26
environment_appropriate	1.01	0.56	545	26
environment_accessible	1.00	0.73	545	26
environment_open	1.00	0.67	545	26
environment_outdoor	0.89	0.71	545	26
relationships_peers	1.29	0.52	519	52
relationships_staff	1.20	0.54	519	52
relationships_acknowledge	1.23	0.54	502	69
relationships_discipline	1.23	0.55	500	71
curriculum_ncf	0.91	0.62	545	26
curriculum_plan	1.00	0.63	545	26
curriculum_balance	1.04	0.58	545	26
curriculum_numeracy	0.94	0.60	545	26
curriculum_literacy	0.96	0.57	545	26
teaching_choice	0.91	0.63	494	77
teaching_engagement	1.15	0.57	507	64
teaching_participation	1.10	0.56	504	67
teaching_questions	0.95	0.61	501	70
teaching_support	1.06	0.58	477	94
assessment_observation	0.87	0.65	545	26
assessment_systematic	0.94	0.63	545	26

There was a fair amount of missing data, and the variable 'teaching_support' in particular has 94 cases missing. This could be a threat to the Factor Analysis, as it means losing 94/571 cases. We need to know more about these missing cases - why are they missing? This is not a question we can answer, but should be addressed[15].

During the Exploratory Factor Analysis reported by Alexander in 2020, 'routine' was found to load poorly in various models. Based on these results and further team discussions on the nature of the item wording, 'routine' was removed from the instrument, and will therefore not be considered here.

1.1. Distribution of individual items

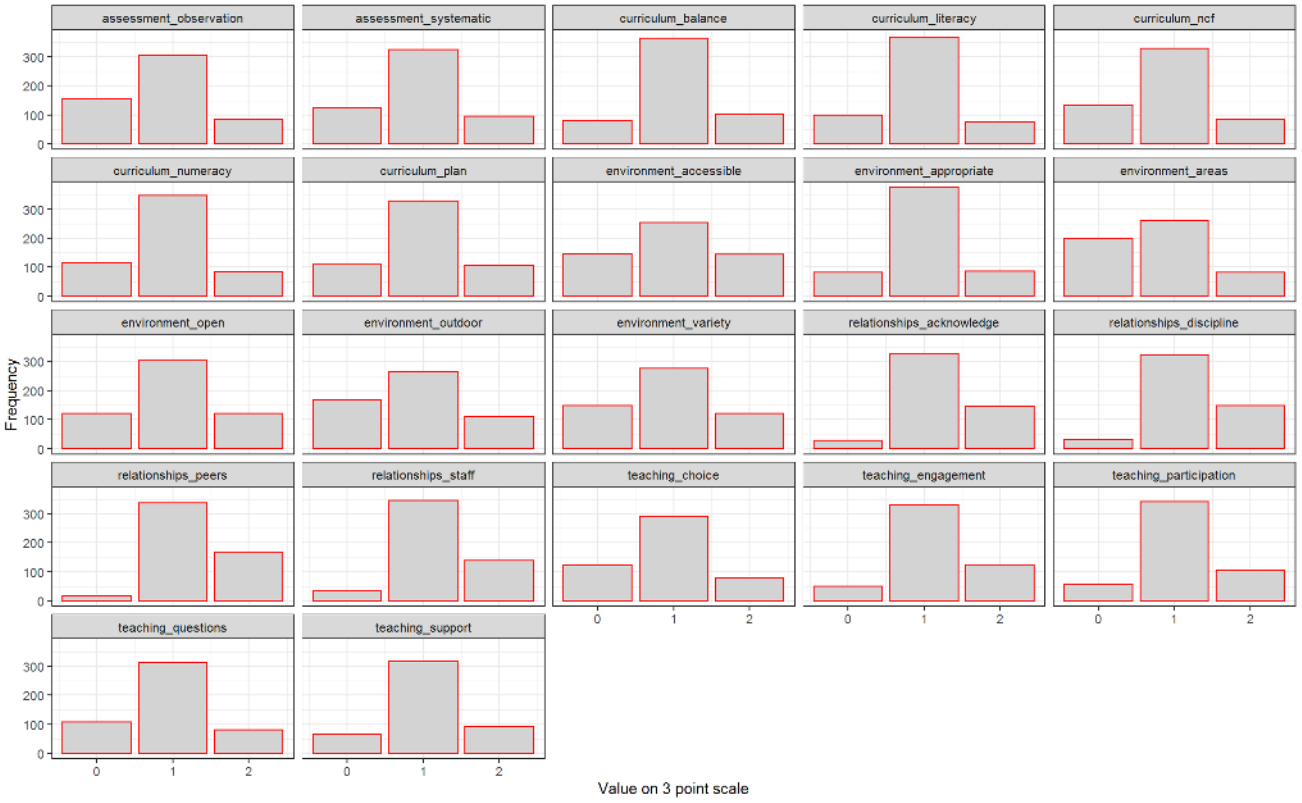


***Figure 1**: Distributions of Items in the LPQA*

Each item in the 22 collected is scored on a 3 point scale. The figure shows that each of the items shows some variation, with a modal response of 1 in all instances

1.2. What type of factor analysis to use?

The analysis reported by Alexander in 2020 was an Exploratory Factor Analysis and she suggested doing Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) as a follow up, with a larger sample size (specifically recommending EFA on a training set, and CFA on a test set).

We do not favour that approach, but it was likely appropriate in 2020 with a small sample size. The scale in question was in fact designed to measure quality on five dimensions, and we prefer to conduct a CFA on the scale, testing the factor and factor-item correspondences proposed in the original scale. There is room to identify items that do not load well, and we can test whether a five factor solution is appropriate. Exploratory Factor Analysis would be appropriate if we were trying to identify reasonable underlying latent variables, but that does not seem to be the case with the LPQA - the underlying factors are in fact clear dimensions of functioning on which ELPs are to be assessed.

We will therefore assess the LPQA through Confirmatory Factor Analysis, assuming the five dimensions outlined above, and with individual items mapped on to factors as specified in the LPQA subscales. We will use the R Programming

[15] Subsequent commentary from reviewers of this report have indicated that "… some items have a 'not observed' check and that might account for some of the missings"

language[16], with packages Lavaan[17] (and to some extent, Psych[18]) to assist us in this task. Where Alexander used Principal Axis Factoring we use Maximum Likelihood Analysis.

One consideration in addition is that Alexander treated individual items in the LPQA as continuous, but there is some question about this, as items only have three values (0, 1, and 2). The items might better be considered ordinal, but that would involve a different approach to the analysis, and could be done on request, but will not be reported here.

A useful way to start any Factor Analysis is by considering the full inter-correlation matrix. This is reported in the Table below.

***Table 2**: Intercorrelations of items in the LPQA*

Item	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	I21	I22
I1	1.00	0.62	0.49	0.54	0.54	0.30	0.17	0.18	0.18	0.18	0.34	0.37	0.37	0.38	0.35	0.33	0.25	0.27	0.28	0.18	0.36	0.28
I2	0.62	1.00	0.58	0.56	0.59	0.36	0.22	0.19	0.27	0.18	0.34	0.36	0.35	0.38	0.37	0.39	0.29	0.23	0.23	0.20	0.31	0.32
I3	0.49	0.58	1.00	0.53	0.51	0.30	0.27	0.27	0.25	0.24	0.38	0.41	0.36	0.41	0.44	0.33	0.24	0.25	0.19	0.26	0.33	0.36
I4	0.54	0.56	0.53	1.00	0.54	0.28	0.26	0.23	0.25	0.17	0.33	0.33	0.31	0.35	0.40	0.37	0.22	0.29	0.29	0.19	0.27	0.29
I5	0.54	0.59	0.51	0.54	1.00	0.29	0.18	0.20	0.21	0.11	0.29	0.32	0.33	0.33	0.37	0.30	0.23	0.25	0.25	0.27	0.25	0.30
I6	0.30	0.36	0.30	0.28	0.29	1.00	0.11	0.16	0.21	0.16	0.34	0.28	0.28	0.24	0.23	0.20	0.16	0.20	0.16	0.16	0.27	0.24
I7	0.17	0.22	0.27	0.26	0.18	0.11	1.00	0.49	0.36	0.37	0.27	0.34	0.24	0.24	0.26	0.30	0.22	0.33	0.33	0.18	0.21	0.26
I8	0.18	0.19	0.27	0.23	0.20	0.16	0.49	1.00	0.39	0.42	0.29	0.27	0.30	0.25	0.35	0.24	0.33	0.38	0.35	0.20	0.22	0.24
I9	0.18	0.27	0.25	0.25	0.21	0.21	0.36	0.39	1.00	0.29	0.26	0.28	0.28	0.26	0.32	0.35	0.30	0.39	0.36	0.22	0.15	0.16
I10	0.18	0.18	0.24	0.17	0.11	0.16	0.37	0.42	0.29	1.00	0.24	0.31	0.29	0.26	0.28	0.21	0.23	0.33	0.23	0.13	0.18	0.20
I11	0.34	0.34	0.38	0.33	0.29	0.34	0.27	0.29	0.26	0.24	1.00	0.52	0.32	0.31	0.38	0.22	0.24	0.28	0.22	0.21	0.36	0.36
I12	0.37	0.36	0.41	0.33	0.32	0.28	0.34	0.27	0.28	0.31	0.52	1.00	0.31	0.38	0.42	0.29	0.29	0.38	0.30	0.20	0.38	0.36
I13	0.37	0.35	0.36	0.31	0.33	0.28	0.24	0.30	0.28	0.29	0.32	0.31	1.00	0.36	0.40	0.37	0.28	0.28	0.27	0.22	0.29	0.29
I14	0.38	0.38	0.41	0.35	0.33	0.24	0.24	0.25	0.26	0.26	0.31	0.38	0.36	1.00	0.53	0.31	0.30	0.32	0.35	0.37	0.32	0.37
I15	0.35	0.37	0.44	0.40	0.37	0.23	0.26	0.35	0.32	0.28	0.38	0.42	0.40	0.53	1.00	0.37	0.34	0.37	0.37	0.32	0.32	0.37
I16	0.33	0.39	0.33	0.37	0.30	0.20	0.30	0.24	0.35	0.21	0.22	0.29	0.37	0.31	0.37	1.00	0.25	0.35	0.25	0.21	0.23	0.19
I17	0.25	0.29	0.24	0.22	0.23	0.16	0.22	0.33	0.30	0.23	0.24	0.29	0.28	0.30	0.34	0.25	1.00	0.35	0.32	0.28	0.24	0.20
I18	0.27	0.23	0.25	0.29	0.25	0.20	0.33	0.38	0.39	0.33	0.28	0.38	0.28	0.32	0.37	0.35	0.35	1.00	0.38	0.37	0.23	0.22
I19	0.28	0.23	0.19	0.29	0.25	0.16	0.33	0.35	0.36	0.23	0.22	0.30	0.27	0.35	0.37	0.25	0.32	0.38	1.00	0.34	0.18	0.28
I20	0.18	0.20	0.26	0.19	0.27	0.16	0.18	0.20	0.22	0.13	0.21	0.20	0.22	0.37	0.32	0.21	0.28	0.37	0.34	1.00	0.18	0.19
I21	0.36	0.31	0.33	0.27	0.25	0.27	0.21	0.22	0.15	0.18	0.36	0.38	0.29	0.32	0.32	0.23	0.24	0.23	0.18	0.18	1.00	0.52
I22	0.28	0.32	0.36	0.29	0.30	0.24	0.26	0.24	0.16	0.20	0.36	0.36	0.29	0.37	0.37	0.19	0.20	0.22	0.28	0.19	0.52	1.00

The correlation matrix is large, and complex, given that there are 22 items in it. The figure below tries to simplify inspection through a gradient colour mapping (a correlogram), where higher correlations are dark blue, and low correlations light blue.

[16] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[17] Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1–36. doi: *10.18637/jss.v048.i02*.
[18] Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois.* R package version 2.2.3. https://CRAN.R-project.org/package=psych.

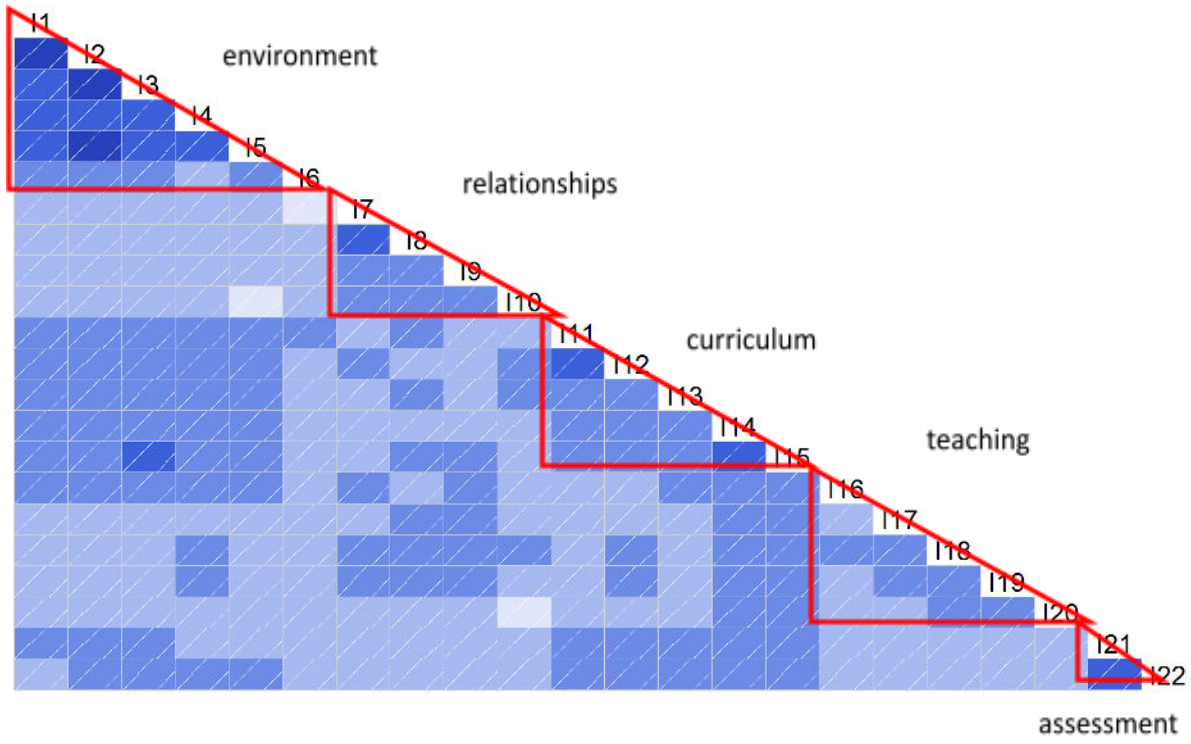


***Figure 2**: Correlogram of Items in the LPQA*

The correlogram above suggests the possibility of factors: quite strongly for environment (I1 - I6), less strongly for relationships (I7 - I10), curriculum (I11 - I15), or teaching (I16 - I20), but strong again for assessment (I21-I22).

1.3. Confirmatory Factor Analysis

The original LPQA items are named according to the factor they are supposed to measure, so the proposed five factor structure is clear by looking at the table of intercorrelations, and as defined earlier in the report.

We start by testing whether a hypothesised five factor solution fits the observed data well. Measures of fit are as follows:

a) The fit, measured by a Chi Squared statistic, against a so-called 'null model', which for our purposes can be considered a model with very little structure in it (that is, there are no-to-few latent variables or domains that underlie the observed data). We would expect the Chi squared value to be significant in this situation.

b) The fit, measured by a Chi Squared statistic, against a so-called 'saturated model', which for our purposes can be considered a model with an excessive amount of structure in it (every item is clearly the product of a specific latent variable or domain). We would prefer the Chi squared value to be *non significant* in this situation, but since Chi squared is sensitive to sample size, and sample size is large in our analysis, it could be significant and yet not reflect poor fit.

c) As indicated above, Chi-squared is sensitive to sample size, and a significant Chi value often does not mean that the data is a poor fit to the model. There are a great many alternative measures of degree of fit that are commonly used, and we will report four here. Firstly, RMSEA (Root Mean Square Error of Approximation), which should ideally by less than .05 for good fit of the model to the data (and the entire 90% confidence interval should preferably be below .05); second, SRMR (Standardized Root Mean Square Residual), which should ideally be smaller than .05; thirdly, TLI (Tucker Lewis index) which should be greater than 0.95, and CFI

(Comparative Fit Index), which should also be greater than 0.95. Extensive discussion of these fit indices, with justifications for the criteria listed above, can be found in Kline (2016[19]).

1.3.1. The five factor model, as originally proposed

***Table 3**: Model fit statistics for original five factor LPQA*

Fit statistic	Value
baseline.chisq	2,742.61
baseline.df	231.00
baseline.pvalue	0.00
Chisq	324.96
Df	199.00
Pvalue	0.00
Cfi	0.95
Tli	0.94
Rmsea	0.04
rmsea.ci.lower	0.03
rmsea.ci.upper	0.05
Srmr	0.04

The table shows that the two Chi squared values are as we expect (both are large, and significant, indicating much better fit than a null model (baseline.chisq), and worse fit than could be achieved with a saturated model (chisq), except that the five factor model fit is larger than we expect (baseline.df should be closer to df = 199). The cfi is quite acceptable, and the tli only slighter low than conventionally accepted. Many researchers would accept the five factor model as fit here.

Individual loadings (both unstandardized and standardized, along with z statistics and p values) are reported below.

[19] Kline, R. B. (2016). Principles and practice of structural equation modelling (4th ed.). Guilford Publications.

***Table 4**: Factor-item loadings for five factor LPQA*

Factor	Item	Unstandardized						Standardized					
		est	Se	z	p <	ci lower	ci upper	est	se	z	p <	ci lower	ci upper
environment	areas	1.00	0.00			1.00	1.00	0.72	0.03	24.77	0.00	0.67	0.78
environment	variety	1.12	0.08	13.85	0.00	0.96	1.28	0.77	0.03	30.20	0.00	0.72	0.82
environment	appropriate	0.79	0.06	12.52	0.00	0.67	0.91	0.70	0.03	22.49	0.00	0.64	0.76
environment	accessible	1.07	0.08	12.83	0.00	0.90	1.23	0.72	0.03	24.02	0.00	0.66	0.77
environment	open	0.94	0.07	12.66	0.00	0.79	1.08	0.71	0.03	23.14	0.00	0.65	0.77
environment	outdoor	0.71	0.08	8.87	0.00	0.55	0.87	0.49	0.04	11.42	0.00	0.41	0.58
relations	peers	1.00	0.00			1.00	1.00	0.62	0.04	14.71	0.00	0.54	0.70
relations	staff	1.06	0.12	8.95	0.00	0.82	1.29	0.64	0.04	15.65	0.00	0.56	0.72
relations	acknowledge	0.98	0.12	8.51	0.00	0.76	1.21	0.60	0.04	13.72	0.00	0.51	0.68
relations	discipline	0.92	0.12	7.81	0.00	0.69	1.15	0.53	0.05	11.34	0.00	0.44	0.62
curric	ncf	1.00	0.00			1.00	1.00	0.57	0.04	14.60	0.00	0.50	0.65
curric	plan	1.11	0.12	9.44	0.00	0.88	1.34	0.64	0.04	17.92	0.00	0.57	0.71
curric	balance	0.90	0.10	8.78	0.00	0.70	1.11	0.57	0.04	14.65	0.00	0.50	0.65
curric	numeracy	1.10	0.11	9.82	0.00	0.88	1.32	0.68	0.03	20.46	0.00	0.61	0.74
curric	literacy	1.03	0.11	9.65	0.00	0.82	1.24	0.66	0.03	19.23	0.00	0.59	0.72
teaching	choice	1.00	0.00			1.00	1.00	0.54	0.04	12.70	0.00	0.46	0.63
teaching	engagement	0.92	0.12	7.71	0.00	0.69	1.15	0.53	0.04	12.08	0.00	0.44	0.61
teaching	participation	0.97	0.12	8.46	0.00	0.75	1.20	0.61	0.04	15.28	0.00	0.53	0.68
teaching	questions	0.94	0.12	7.89	0.00	0.71	1.18	0.55	0.04	12.74	0.00	0.46	0.63
teaching	support	0.84	0.12	7.29	0.00	0.61	1.07	0.49	0.05	10.68	0.00	0.40	0.58
assess	observation	1.00	0.00			1.00	1.00	0.72	0.04	17.40	0.00	0.64	0.80
assess	systematic	0.99	0.10	9.57	0.00	0.79	1.19	0.74	0.04	18.13	0.00	0.66	0.82

est = estimate (coefficient), se = standard error, z = z statistic, p = probability value

The fit indices referred to above are of global fit (i.e., averaged over the whole model), so we need to explore local fit issues to see where problems and lack of fit might be.

Inspection of the standardized residuals[20] and modification indices[21] extracted from the model fit (shown in the Appendices) showed that only two items appeared to have poor local fit, namely the items 'ncf' and 'plan', in the Curriculm subscale. We decided to allow correlated error terms between these variables, which inspection of the modification indices suggested would be beneficial. Correlating indicator error terms can improve model goodness of fit and improve the reliability of the latent construct's scale. This is not at all unusual when examining problems of local fit.

Inspection of the standardized loadings also suggested that the items 'tdr' (outdoor) and 'spprt' (support) on the Environment and Teaching scales had lower loadings than desired (< .50). In the interest of shortening the LPQA these items could be dropped from those subscales. If we do that, we get the following solution for a revised five factor model.

***Table 5**: Model fit statistics for revised five factor LPQA*

Fit statistic	Value

[20] Residuals are the scores representing the differences between the observed correlations and those predicted by the model. These are standardized so as to be interpretable as standard normal deviates, allowing computation of probability values

[21] Modification indices are the amount that the model Chi squared will change with inclusion or exclusion of model components.

baseline.chisq	2,641.49
baseline.df	190.00
baseline.pvalue	0.00
Chisq	236.36
Df	159.00
Pvalue	0.00
Cfi	0.97
Tli	0.96
Rmsea	0.04
rmsea.ci.lower	0.03
rmsea.ci.upper	0.04
Srmr	0.04

The following table shows that revised five factor model fits the data well at a global level (the model Chi squared is lower than for the original five factor model, although it is still statistically significant, indicating that a better fit is possible). All the fit indices are well within values considered a very good fit by most researchers.

Individual loadings (both unstandardized and standardized, along with z statistics and p values) are reported in the following table.

***Table 6**: Factor-item loadings for revised five factor LPQA*

Factor	Item	Unstandardized						Standardized					
		est	Se	z	p <	ci lower	ci upper	est	se	z	p <	ci lower	ci upper
environment	areas	1.00	0.00			1.00	1.00	0.73	0.03	25.60	0.00	0.67	0.78
environment	variety	1.11	0.08	14.22	0.00	0.96	1.26	0.78	0.03	30.73	0.00	0.73	0.83
environment	appropriate	0.77	0.06	12.73	0.00	0.65	0.88	0.69	0.03	22.41	0.00	0.63	0.75
environment	accessible	1.05	0.08	13.26	0.00	0.90	1.21	0.72	0.03	24.90	0.00	0.66	0.78
environment	open	0.93	0.07	13.11	0.00	0.79	1.07	0.71	0.03	24.14	0.00	0.65	0.77
relations	peers	1.00	0.00			1.00	1.00	0.61	0.04	14.70	0.00	0.53	0.69
relations	staff	1.06	0.12	9.01	0.00	0.83	1.29	0.63	0.04	15.62	0.00	0.55	0.71
relations	acknowledge	1.04	0.12	8.78	0.00	0.81	1.27	0.60	0.04	14.55	0.00	0.52	0.69
relations	discipline	0.93	0.12	7.94	0.00	0.70	1.15	0.53	0.05	11.62	0.00	0.44	0.61
curric	ncf	1.00	0.00			1.00	1.00	0.55	0.04	13.67	0.00	0.47	0.63
curric	plan	1.11	0.10	10.73	0.00	0.91	1.32	0.61	0.04	16.66	0.00	0.54	0.68
curric	balance	0.96	0.11	8.80	0.00	0.75	1.18	0.59	0.04	15.54	0.00	0.51	0.66
curric	numeracy	1.13	0.12	9.55	0.00	0.90	1.36	0.67	0.03	20.19	0.00	0.60	0.73
curric	literacy	1.08	0.11	9.50	0.00	0.85	1.30	0.66	0.03	19.83	0.00	0.60	0.73
teaching	choice	1.00	0.00			1.00	1.00	0.55	0.04	13.36	0.00	0.47	0.63
teaching	engagement	0.89	0.11	7.99	0.00	0.67	1.11	0.51	0.04	11.99	0.00	0.43	0.60
teaching	participation	0.94	0.11	8.74	0.00	0.73	1.15	0.59	0.04	14.72	0.00	0.51	0.66
teaching	questions	0.89	0.11	8.08	0.00	0.67	1.11	0.52	0.04	12.27	0.00	0.44	0.61
assess	observation	1.00	0.00			1.00	1.00	0.71	0.04	17.50	0.00	0.63	0.79
assess	systematic	0.99	0.10	9.66	0.00	0.79	1.19	0.75	0.04	18.47	0.00	0.67	0.83

est = estimate (coefficient), se = standard error, z = z statistic, p = probability value

There seems to be no good reason to modify the scales any further. The global fit is very good, and there do not seem to be any clear issues of problems with local fit.

If the scales are used as is, they have reasonably good internal consistency values, bearing in mind that the scales are brief, sometimes having as few as two items.

***Table 7**: Omega and average inter item correlation for revised five factor LPQA*

Statistic	Environment	Relationships	Curriculum	Teaching	Assessment
Omega	0.86	0.72	0.77	0.65	0.68
Ave. inter-item r	0.55	0.39	0.39	0.32	0.52

Of the five scales, Teaching and Assessment both have reliability indices (Omega) lower than would commonly be accepted as suitable (0.7), but the Assessment scale only has two items, and a high intercorrelation between them (r = 0.52), so can be judged acceptable. The Teaching subscale has a slightly lower index than typically recommended, but the average inter-item correlation is 0.32, and this is higher than a commonly accepted minimum value of r = 0.25[22]

Although internal consistency and factorial validity are important criteria for assessing scales, it is perhaps more important to consider evidence that the scales are correlated with outcomes that they ought to be correlated with, from a theoretical or even common-sense point of view. We therefore investigated the correlations between the scales and three measures of socio-economic status (whether an ELP received subsidy, school quintile, and the (log-transformed) average fee charged for children in the target age range of 50 to 50 months, as well as the key outcome measure of the Thrive by Five index, the total ELOM 4-5 score). Since the scales are intended to assess ELP quality, and we know that socio-economic advantage allows ELPs to provide better quality service, at least some of the scales ought to correlate with the socio-economic measures. On the other hand, assuming that better quality provisioning of ELPs is related to final outcome i.e., higher total ELOM 4-5 score, some of the scales ought to be correlated with the outcome.

We therefore created unit-weighted total scores for each of the five scales, and computed correlations, and scatterplots of relations. Note that we created unit-weighted total scores, in which items in each subscale are summed, rather than total scores weighted by factor loadings. The correlations we report would likely be higher if scales were composed by weighting items according to loadings, but it is a good test of robustness to use unit-weighted scales, as this is the most common way practitioners end up using them.

The following figure reports a scatterplot and correlation matrix:

[22] We revisited the fit of the revised five factor model and tested specifically whether it was any better than a four factor model that excluded the Teaching subscale. In fact, the four factor model was slightly preferable to the five factor model, a LR Chi Square test yielding Chi squared = 82.6, df = 62, $p < .041$. However, one should bear in mind that the sample on which the data was collected was large, making Chi squared tests very sensitive, and we do not make much of this difference.

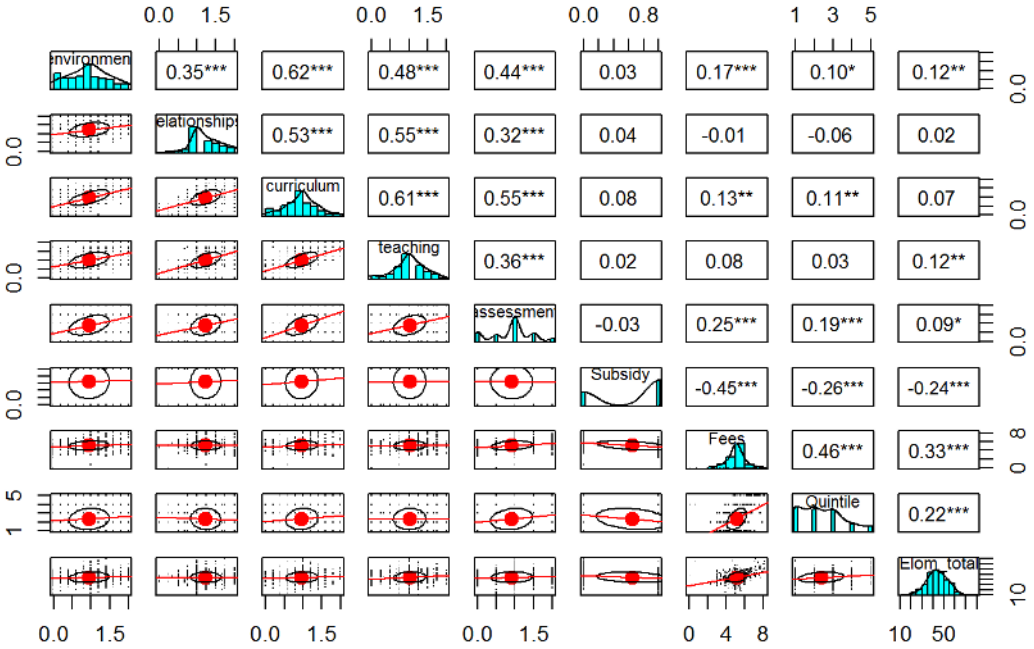


***Figure 3**: Scatterplot matrix of relations between scale scores and criterion variables*

The figure in question shows that the subscale measuring Environment correlates significantly with socio-eceonomic measures (subsidy, quintile, fees), and with the total ELOM 4-5 score. The same is not true for the Relationships scale. The Curriculum scale correlates with socioeconomic variables, but not with total ELOM 4-5. The Teaching scale does not correlate with socioeconomic variables, but does correlate with total ELOM 4-5 performance. The Assessment scale, like the Environment scale, correlates with both socioeconomic variables and the total ELOM 4-5 scale. The overall conclusion is that four of the five scales correlate with criterion variables as we expected them to. The only scale that did not do so is that measuring relationships, and it might well be appropriate that it did not correlate with the criterion variables, given the nature of the items in the scale.

The recommendation is thus to keep the five subscales as originally devised, with minor modifications, as detailed earlier in the report. It is also possible to keep the five subscales as originally devised, with no modifications, accepting that this yields fit indices that are acceptable on most but not all of the criteria conventionally accepted in the literature.