Measurement Issues in Criterion Q-sorts: the Assessment of Social Competence

2 August 2022

Abstract

Measurement issues are foundational for our understanding of human behaviour. This paper discusses such issues for criterion Q-sorts, focusing on the criterion Q-sort for social competence developed by Waters, Noyes, Vaughn and Ricks (1985). These issues appear to have been overlooked, with the result that this criterion Q-sort has been misused (e.g., by Vaughn, Shin, et al, 2009 and Vaughn, Santos, et al., 2016), affecting our understanding of social competence. To illustrate the pragmatic importance of these issues, we present strong replicated differences in relations between two measures of social competence (Waters'criterion Q-sort and an unidimensional Q-scale) and measures of children's bullying, prosocial behaviour, and cooperativeness. Because correlations measure distance but not direction, individual criterion scores only indicate degree of similarity with a criterion but provide no information on *how* the individual differs from the criterion – a limitation, or characteristic, that needs to be kept in mind as we seek to identify and understand varieties of social competence.

Keywords: Criterion Q-sorts; measurement issues; social competence; aggression; prosocial behaviour.

Public significance: It has generally not been recognized that because a criterion Q-sort specifies the relations between its target construct (such as social competence) and all other constructs assessed by the Q-set, that it cannot be used to investigate relations between that target construct and those other constructs (for example, children's aggression or prosocial behavior). These relations have already been specified. Nor can it be used in an exploratory fashion to identify variants of its target construct (such as different ways in which children may be socially competent in different contexts) since correlations index distance but not direction. Recognition

of these issues should improve our ability to investigate and understand social competence in children.

Measurement Issues in Criterion Q-sorts: the Assessment of Social Competence

Measurement issues are foundational for our understanding of human behaviour. This paper discusses such issues for criterion Q-sorts, focusing on the criterion Q-sort for social competence developed by Waters, Noyes, Vaughn and Ricks (1985). Although obvious, these issues appear to have been overlooked, with the result that this important instrument has been misused in research that attempts to identify and understand different varieties of social competence that occur in different contexts (e.g., Vaughn, Santos, et al., 2016 and Vaughn, Shin, et al, 2009). To illustrate the practical importance of these issues, we present strong replicated differences in relations between two measures of social competence (Waters' multidimensional criterion Q-sort and a unidimensional Q-scale) and unidimensional measures of other important social behaviours: children's bullying, prosocial behaviour, and cooperativeness.

Q-sorts are powerful and useful tools for describing behavior. In comparison with questionnaire data, Q-data are more thoughtful and meaningful because of the constraints placed on sorting the items. It is an important strength of Q-sorts that they can directly assess behaviours and beliefs across diverse domains, allowing researchers to gain a view of the relative importance of diverse behaviors (Block, 2008; Calder-Stegemann & Roberts, 2016). For example, the 100 items of the California Child Q-set (Block, 2008; Block & Block, 1969) describe over 30 aspects of behaviour (Block & Block, 1980, Table 2.6), and the 72 items of the Preschool Behavior Q-set (Baumrind, 1968), are nearly as diverse (Baumrind, 1971; see Table 2, which presents only 45 items from the Q-set). These are the two Q-sets for which Waters and his colleagues developed criterion Q-sorts for social competence.

Criterion Q-sorts define a target construct by specifying the rank of every item in the Q-set (see Table 1 in Waters, Noyes, Vaughn & Ricks, 1985, for examples). By so doing, they define relations between *all* constructs assessed by the Q-set. Thus constructing a criterion Q-sort requires broad, integrative theory, since it represents a consensus not only about the specific behaviours that define a target construct (such as social competence), but of how those behaviours are related to the other constructs assessed by the Q-set. They thus serve an important role in clarifying and developing theory.

Criterion Q-sorts can also be used to assess individuals. This is done by correlating their Q-sorts with the criterion Q-sort (Block, 2008). Such a correlation can be used as an index of social competence (e.g., Vaughn, Santos, Monteiro, Shin, Daniel, & Krzysik, 2016). In this process, the measurement of the target construct and all other constructs in the Q-set become summarized in a single number.

Although never desirable, this confounding is perhaps not a serious problem for conceptually broad criterion Q-sorts, ones whose target construct involves all or nearly all the items in the Q-set. Examples would include criterion Q-sorts for optimal mental health in adults (Block, 2008), general competence in preschool children (Roberts & Strayer, 1987), and overall excellence in teaching (Calder-Stegemean & Roberts, 2016). In these cases, it is obvious that the correlation is a summary incorporating diverse domains.

However, the case is quite different for conceptually narrow criterion Q-sorts, ones that target a specific construct in a diverse, multidimensional Q-set. The criterion Q-sort for social competence developed by Waters et al. (1985) is of this type (although this was not the understanding of the authors). In such a case, as we will illustrate below, correlations with the

criterion Q-sort can be strongly influenced by non-target constructs. This obvious problem apparently has not been recognized in the Q-sort research literature, even though individual assessment using Waters' criterion Q-sorts is not uncommon (e.g., Vaughn et al., 2016). Indeed, such correlations can now be calculated using R (Sousa & Daniel, 2018).

In a conceptually narrow criterion Q-sort, distortions arise because the target construct is defined by relatively few items in the Q-set. In such a case, non-target items must supplement target items at the extremes of the (forced, or specified) Q-distribution. As readers are aware, extreme items have a strong influence on the value of a correlation. This occurs because correlations are the average cross-product of *z*-scores, and so extreme items (which have large z-scores) have greater influence on the magnitude of correlations than items near the mean (which have *z*-scores near zero). For example, in the case of the Preschool Behaviour Q-set (72 items distributed equally across nine categories from *least* to *most characteristic*), the eight items with a rank of 5 have z-scores of zero and do not influence the sum of cross-products at all. In contrast, the 16 items in categories 1 and 2 have *z*-scores of -1.5 and -1.2, and the 16 items in categories 8 and 9 have z-scores of +1.2 and +1.5. Thus in a criterion sort based on the Preschool Q-set, the extreme 32 items will most strongly influence the magnitude of its correlation with other Q-sorts. A similar pattern holds for the California Child Q-set, with its quasi-normal distribution.

Obviously, the fewer the number of target-specific items to place at the extremes, the more opportunity for other constructs to distort the correlation with the criterion as a measure of the target construct. And in general, target-specific items are few, regardless of the construct, because a Q-set that assesses many dimensions of behaviour can do so only by limiting the

number of items relevant to each construct (J. H. Block, 1965). The second paragraph of this paper suggests an average of six or seven items per construct for the Preschool Q-set and three or four items for the California Child Q-set. Thus a criterion Q-sort that focuses on one construct (such as social competence) in a multi-dimensional Q-set will be unable to monopolize the most extreme positions in the criterion distribution. Other constructs will be there, too. An examination of the 32 most-extreme items in Waters' criterion sort for social competence for the Preschool Behaviour Q-set (Table 1 in Waters et al., 1985) indicates that only about six reflect the core meaning of social competence as it is now usually understood: engagement with, and acceptance by, peers.

Obvious as this problem now is, it was not apparent in the 1980s when Waters' criterion Q-sort was developed. At that time, social competence was considered to be a conceptually broad construct. Prosocial behaviour with peers, cooperative behaviour with adults, and aggressive, bullying behaviours were then not considered to be correlates of social competence (Rose-Krasnor, 1997): They were considered to be aspects of social competence. Social competence *by definition* included prosocial behaviours, cooperation, emotional regulation, purposiveness, dominance, and non-aggressiveness (e.g., Crick & Dodge, 1999; Rose-Krasnor, 1997; Sroufe, Egeland, Carlson, & Collins, 2005; Waters & Sroufe, 1983). These views were naturally reflected in Waters' criterion Q-sort. There was no difficulty in finding relevant items to fill the extremes of the criterion distribution.

Views of social competence shifted during the1990s, becoming more differentiated. For example, observational studies found that aggressive children were sometimes socially skilled and even prosocial (e.g., Pepler, Craig, & Roberts, 1995; Sutton, Smith, & Swettenham, 1999;

Vaughn, Vollenweider, Bost, & Azria-Evans, 2003). Similarly, it was found that children who were not aggressive in the classroom were often aggressive on the playground (e.g., Pepler, Craig, & Roberts, 1996, 1998). Thus children considered by teachers to be socially competent were sometimes aggressive, and children designated as aggressive often displayed social skills. These and similar findings led Hawley & Vaughn (2003) and Bukowski (2003) to argue that moderate levels of aggressiveness enhanced dominance and prestige and therefore were often attractive to peers. It is now thought that moderate levels of aggression are compatible with peer acceptance in younger children (e.g., Ladd, 2005; Hawley, 2006), although certainly not necessary.

These views imply that there are diverse relations between peer engagement and acceptance, on the one hand, and prosocial behaviour and aggression on the other – that there are different ways to be socially competent and different types of socially competent children (e.g., Hawley, 2007; Roberts, 2020; Vaughn et al., 2016). In contrast to these differentiated views, a criterion Q-sort can assess only one type of social competence, the particular type it defines. In the case of Waters' criterion Q-sort, this is peer engagement and acceptance accompanied by friendly, prosocial behaviour and very little aggression. It is obvious that because Waters' criterion Q-sort specifies these relations, it cannot be used to investigate them – or to investigate relations between social competence and any other aspect of behaviour assessed by the Q-set. In addition, correlations measure distance but not direction. Thus individual criterion scores only indicate degree of similarity with a criterion. They provide no information on *how* individuals differ from the criterion. Thus a criterion Q-sort cannot be used to identify other types of social competence. These limitations, or characteristics, need to be kept in mind

These comments, *mutatis mutandis*, apply to all criterion Q-sorts focused on a characteristic that comprises only one facet of a multifactorial Q-set. Thus they apply to criterion Q-sorts for the California Child Q-Set, for example, as well as to those based on the Preschool Behavior Q-set. Obvious as it is, this issue apparently has been unacknowledged in the Q-sort research literature, as noted above. But it is important that it be kept in mind.

That criterion Q-sort scores can be strongly influenced by other constructs in the Q-set will be illustrated by analyzing Q-data from two earlier community samples of preschool children, one from Vancouver, Canada (Roberts & Strayer, 1987) and the other from Toronto (Roberts, 2020). These analyses contrast relations between three important aspects of children's social behaviour (bullying, prosocial behavior with peers, and cooperative behaviour with adults) and two measures of social competence: scores on the criterion Q-sort for social competence (Waters et al., 1985) and scores on a conceptually simple, unidimensional Q-scale of peer social competence.

Analyzing two samples allows us to assess whether patterns of distortion replicate. The need for replication is often ignored and sometimes even discounted in psychology (e.g., Rovenpor & Gonzales, 2015). But it is the gold standard for scientific findings (Chambers, 2014, June 10). Among other things, replication directly refutes concerns about false positives in small samples (e.g., Cohen, 1994). They cannot be dismissed, especially when the replicating samples are small (as they are here). Because sampling error is greater in small samples, replication is more difficult. Thus when samples are small, replication suggests that findings are unusually robust and dependable – just as the arguments above suggest they should be.

Because the analyses below were done within samples and were based on the same

instrument (the Preschool Behavior Q-set), sample characteristics are controlled: They are the same for all comparisons. Therefore the paradoxical findings presented below must arise from the measures themselves, for the reasons discussed above, not from characteristics of the samples.

Nevertheless, it will be useful to begin by briefly describing sample characteristics and summarizing the methods used to collect the data. We will then describe the Q-scales in detail and compare scale scores with scores from the social competence criterion Q-sort.

The Empirical Examples

The Samples

Vancouver. As reported in Roberts (1987) as well as Roberts and Strayer (1987), the ages of the 30 children (63% female) who participated in this study ranged from 3.0 to 5.8 years, mean = 4.3 years. Over two-thirds (21) had at least one sibling. Fathers' average age was 34 years; mothers', 32. Mothers reported an average of 14 years of schooling; fathers, 16 (range for both, 9 to 21). Mean family income was somewhat above the national average. Duncan SEI scores ranged from 11 to 92, mean = 59.

Each child's daycare or preschool teacher completed the Preschool Behavior Q-set (Baumrind, 1968), which required them to evenly distribute 72 items across nine categories from *extremely uncharacteristic* (= 1) to *extremely characteristic* (= 9) of the child. For five cases, Q-sorts were available from two teachers. Their average correlation was .69, a value similar to that reported by Baumrind (1971).

Family interactions were assessed by home observations (focal-individual samples of behaviour from supper time to the child's bedtime, on an evening when both parents were at

home), observer ratings, parent self-reports (the Child-Rearing Practices Q-set, Block, 1965) and a child interview. Further details of measures and methods are reported in Roberts and Strayer (1987).

Toronto. As reported in Roberts (2020), the ages of the 33 children (48% female) who participated in this study ranged from 4.1 to 6.1 years, mean = 4.8 years. Nearly three-quarters (73%) had at least one sibling. Fathers' mean age was 38 years; mothers', 35. Mothers and fathers each reported an average of 16.3 years of schooling (range 10 to 26). Median family income was $98,000 in 2015 Canadian dollars (range = $17,000–$237,00). In comparison to Vancouver families, Toronto families were more ethnically diverse, with only 85% speaking English at home (other languages included Spanish, Italian, Arabic, and Punjabi).

As in the Vancouver sample, daycare or preschool teachers completed the Preschool Behavior Q-set (Baumrind, 1968) for study children. In addition, observers collected focal-individual samples of peer behaviour for the study child (to assess social competence) and nearest-neighbour scan samples for the entire class (to assess social networks, i.e., friends). Data were collected during play periods when children were free to move about. Means of 135 min of focal-individual data and 29 scan samples were obtained for each child over an average of 4.4 days. After this, observers completed a Preschool Behaviour Q-set for the study child. Correlations indicated acceptable agreement between observer and teacher Q-sorts, mean r = .50. Following Block (2008), items were aggregated by averaging and then combined to form the scales described below.

Family interactions were assessed by home observations (focal-individual samples of behaviour over a mean of 3.6 evenings when both parents were at home) and parent self-reports

(the Child-rearing Practices Q-set, Block, 1965, as well as questionnaire measures of stress and social support). Further details are given in Roberts and Briner (2015) as well as Roberts (2020).

Q-scales of Social Behaviours

For all scales, dimensionality was assessed by maximum likelihood factor analysis with single-factor solutions indicated by parallel analysis, as recommended by Osborne (2014). Scores were calculated by averaging (Roberts, 2017), after reflecting items that loaded negatively.

Peer Competence. Derived from Waters, Wippman, and Sroufe (1979), the nine items that comprised this scale (*Paid attention to by other children*; *Other children seek his/her company*; *Peer leader*; *Outgoing, frequently interacts with peers and teachers*; *Bold with other children*; *Enjoys excitement*; *Participates in play*; *Suggests activities to other children, initiates play*; *Participates fully in group talks*) focused on active peer interactions and peer acceptance, the core aspects of social competence (e.g., Rose-Krasnor, 1997; Rubin et al., 1998). Cronbach alphas for the Peer Competence scale were .95 in Roberts (2020) and .85 in a re-analysis of Roberts & Strayer (1987).

Bullying. Derived from Baumrind (1971), as were the two scales below, these six items assessed aggressive, bullying behaviour: *Bullies other children*; *Insults, calls names, or ridicules other children*; *Hits other children frequently*; *Domineering, intrusive, impolite*; *Lacks ability to get along with other children*; *Supports or incites misbehavior by other children.* Cronbach alphas were .89 in Roberts (2020) and .70 in a re-analysis of Roberts & Strayer (1987).

Prosocial with peers. Six items assessed behaviours that were friendly, helpful, and kind, but not clearly tied to social engagement as such: *Sympathetic towards other children who are upset*; *Considerate, empathic; Shares possessions willingly*; *Easily includes others in play*; *Helps*

other children carry out their plans; *Careful of others' work*. Cronbach alphas were .84 in Roberts (2020) and .77 in a re-analysis of Roberts & Strayer (1987).

Cooperative with Adults. Six items assessed behaviour that was compliant and cooperative with adults: *Obedient*; *Actively facilitates nursery school routine*; *Accepts adult guidance*; *Does not test limits set by adults*; *Responsible about following standard school procedures*; *Can be trusted*. Cronbach alphas were .94 in Roberts (2020) and .92 in a re-analysis of Roberts & Strayer (1987).

Q-sort Scales and Social Competence Criterion Scores

Convergence in measures of social competence. As expected, scores on both measures of social competence (the Peer Competence Q-scale and correlations with the social competence criterion) were strongly and positively correlated in both samples (*r*s = .52 and .60; Table 1). Nevertheless, they showed different relations with bullying, prosocial behaviour, and cooperativeness, as shown below.

Criterion scores and social behaviours. As shown in Table 1, there were, as expected, substantial negative correlations between criterion scores and bullying in both samples (*r*s = -.43 and -.53), strong positive correlations between criterion scores and prosocial behaviour (*r*s = .62 and .67), and moderately strong positive correlations between criterion scores and cooperative behaviour with adults (*r*s = .37 and .44). These correlations are consistent with items in the extreme ranks (1, 2, 8, and 9) in the criterion Q-sort (Waters et al., 1985, Table 1). Indeed, many Q-scale items are among those extreme items, indicating that the correlations are artifacts of the measures.

Q-scale peer competence and aggression. In contrast to criterion scores and consistent

with the view that peer competent children can be aggressive (e.g., Hawley, 2007), peer competence Q-scale scores were moderately and *positively* correlated with bullying in both samples (*r*s = .34 and .38). These correlations (bullying with Q-scale peer competence and bullying with criterion Q-sort scores) differed significantly and strongly in both samples (see Table 1).

Q-scale peer competence and prosocial behaviour. In contrast to criterion Q-scores, Q-scale peer competence was *unrelated* to prosocial behaviour with peers in both samples (*r*s = .02 and -.10), suggesting that it is not necessary to be kind to be liked, as Waters and his colleagues thought. As with bullying, correlations between prosocial behaviour and the two measures of peer competence differed significantly and strongly in both samples.

Q-scale peer competence and cooperative behaviour. In contrast to criterion scores, peer competence Q-scale scores had moderately strong *negative* correlations in both samples with cooperative behaviour with adults, *r*s = -.29 and -.42. This suggests that non-compliance with adults, like aggression with peers, is compatible with peer engagement and peer acceptance. Q-scale and Q-criterion correlations with cooperativeness differed significantly and strongly in both samples, as shown in Table 1.

Taken together, these findings indicate that results from a variable-centered analysis based on unidimensional Q-scales can differ in conceptually important ways from results based on the social competence criterion Q-sort.

That important conceptual problems can also arise in a typological approach will now be illustrated using the three groups of children described in Roberts (2020) and also found in Roberts and Strayer (1987).

Problems in a Typological Analysis

As reviewed earlier, there are sound theoretical reasons for expecting to find the three groups identified in Roberts (2020) by a km cluster analysis – (i) socially engaged, peer-accepted children who are prosocial (as Waters expected); (ii) socially engaged, peer-accepted children who are aggressive (as Pepler and others have found); and (iii) children who are inactive socially and not salient for their peers. These groups were replicated in a km cluster analysis of data from Roberts and Strayer (1987). Although clearly identified in both samples by unidimensional Q-scale scores for peer competence and bullying, social competence criterion scores distorted the characteristics of the aggressive group and failed to distinguish between groups.

To begin with the distortion: In both samples aggressive children were among those who were most engaged with and most accepted by peers (in the Vancouver sample, their mean score was 6.7 out of 9 on the peer competence scale. For comparison, the mean score for the socially engaged prosocial group was 6.6. In the Toronto sample, mean scores were 7.3 and 7.6 for these two groups, respectively). But despite high levels of peer engagement and acceptance, criterion scores for aggressive children were low (in Vancouver, aggressive children had a mean correlation of .20 with the social competence criterion Q-sort; in Toronto, the mean correlation was .18). Thus in both samples, high levels of peer engagement and acceptance were masked by low criterion scores. (In contrast, the socially active prosocial group, although not more socially engaged, had substantially higher criterion scores in both samples, with means of .56 in Vancouver and .69 in Toronto.)

The reason for this distortion is not far to seek: items for peer competence and bullying occupy opposite extremes of the criterion Q-sort distribution (Waters et al., 1985, Table 1). Thus

in the aggressive group, high positive *z*-scores for items assessing peer engagement and acceptance are cancelled by high negative *z*-scores for items assessing aggression, resulting in a low average *z*-score (correlation). Thus the multidimensional nature of the criterion Q-sort can falsely suggest that aggressive children are *not* characteristically socially engaged and accepted by their peers.

For this same reason – multidimensionality – criterion scores failed to distinguish children who were characteristically aggressive and socially engaged from children who were characteristically neither. In Toronto, the latter group had a mean correlation with the social competence criterion of .31 (vs .18 for the high-aggression group, as just mentioned). These values suggest a substantial overlap of scores between the two groups; and a *t*-test based on group membership accounted for less than 6% of the variance in criterion correlations, $t(20) = 1.08$. Thus despite substantial differences in behaviour (for aggression, mean scale scores were 6.2 out of 9 vs 2.4; for peer competence, 7.3 vs. 4.6), these two groups could not be differentiated by criterion scores.

The same pattern was seen in Vancouver. Correlations with Waters' criterion Q-sort were again similar for both groups (means = .24 and .20, respectively, for the low engagement, low aggression group and the socially engaged, high aggression group). Group membership accounted for less than 1% of the variance in criterion correlations, $t(12) = 0.24$. In contrast, behavioural differences as assessed by unidimensional scales were again substantial. For the aggressive group, both aggression and peer engagement were characteristic, with mean scores of 5.8 and 6.7, respectively; whereas for the low group, aggression and peer engagement were *un*characteristic, with means = 2.8 and 3.8, respectively. Thus in this sample, too, similar

criterion scores masked substantial differences in behaviour.

Discussion

The empirical examples just presented support the basic concern raised by this article: Because criterion Q-sorts involve all items in a Q-set, they define relations between all constructs assessed by the Q-set. Correlations with criterion scores confound these constructs, making it impossible to separate effects or interpret scores. As with all multidimensional measures, there is only one way to achieve a very high correlation with a criterion Q-sort, but many ways to achieve moderate or low correlations. As illustrated above, children with similar criterion scores can differ in important ways in their behaviour, and children with similar levels of a target behaviour can have very different criterion scores.

As noted above, relations with other variables are distorted because correlations with the criterion Q-sort are influenced not only by items assessing the target construct but by extreme items assessing other constructs. There are typically many such items, given that multi-dimensional Q-sets (like the Preschool Behavior and California Child Q-sets) assess many constructs, limiting the number of items relevant to any given one. Although these problems are obvious, they apparently have been ignored in research using Q-methodology.

Many researchers (e.g., Bukowski, 2003; Vaughn et al., 2009, 2016) expect complex relations between peer competence and aggression; that is, they expect to find more types of peer competent children than the three discussed here and in Roberts (2020). Vaughn et al. (2016), for example, identified five types. But in a typological approach, each dimension of a multidimensional space needs to be defined by a unidimensional axis. Thus a global measure that defines a single type of social competence, as Waters' criterion Q-sort does, is unsuitable, even

as part of a battery of measures (e.g., Vaughn et al., 2009). Correlations indicate distance, not direction. They cannot separate groups that are very different but equally far from a core construct.

It is worth emphasizing that a group of children of theoretical and practical significance – children who are both aggressive and engaged with (and accepted by) peers – *cannot* be identified using the social competence criterion Q-sort. Because the criterion Q-sort *defines* peer acceptance and aggression as antagonistic characteristics, these items, as noted above, are placed at opposite extremes of the criterion distribution and cancel each other when children are both aggressive and engaged with peers. The resulting low correlations are similar to criterion correlations for children low on social engagement. The result is that these two groups, as we saw above, cannot be distinguished. In contrast, these groups are easily identified in a multidimensional space defined by unidimensional Q-scales for peer competence and aggression.

It should be obvious that these problems cannot be mitigated by updating or revising Waters' original social competence criterion Q-sort or by developing new criterion Q-sorts to reflect current theory. An up-dated or different criterion Q-sort would simply present a different set of distortions, based on the single type of social competence that it defined. All the constructs assessed by the Q-set, including the target construct, would still be confounded, making it useless for many research purposes.

If, as Vaughn and others have argued, the type of social competence defined by Waters' criterion Q-sort is of prime importance because it is associated with favorable long-term outcomes for children, then it is even more important to identify the particular ways in which children deviate from this type. Scores on Waters' criterion Q-sort cannot do this.

In conclusion, the problems identified in this paper are not specific to the social competence criterion Q-sort. They are inherent in the nature of Q-sets and criterion Q-sorts. As noted earlier, Q-sets usually assess a broad array of characteristics and therefore can devote relatively few items to each. Criterion Q-sorts necessarily define relations between all constructs in a Q-set, and in doing so, confound these constructs and distort the measurement of target constructs when these are defined by relatively few Q-items. These characteristics, and limitations, need to be borne in mind when undertaking and evaluating research that uses criterion Q-sorts.

References

Baumrind, D. (1968). *Manual for the preschool behavior Q-Sort*. Institute of Human Development, University of California, Berkeley.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology Monographs, 4*, (1, Pt.2), 1-103.

Block, J. (2008). *The Q-Sort in character appraisal*. Washington, DC: APA.

Block, J. H. (1965). The child-rearing practices report (CRPR): a set of Q items for the description of parental socialization attitudes and values. Unpublished manuscript, Institute of Human Development, University of California, Berkeley.

Block, J. H. & Block, J. (1969). *The California Child Q-Set*. Institute of Human Development, University of California, Berkeley.

Block, J., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. Collins (Ed.), *Development of cognition, affect, and social relations. Minnesota symposia on child psychology* (Vol. 13, pp. 39-102). Hillsdale: Erlbaum.

Block, J., & Robins, R. (1993). A longitudinal study of consistency and change in self-esteem from early adolescence to early adulthood. *Child Development*, 94, 909–923.

Bukowski, W. (2003). What Does It Mean to Say That Aggressive Children Are Competent or Incompetent? *Merrill-Palmer Quarterly*, 49, 390–400. DOI: 10.1353/mpq.2003.0011

Calder Stegemann, K. & Roberts, W. (2016): Teaching Across the Grades: Dispositions of Ideally Competent Teachers. *The New Educator*, 1-17. DOI: 10.1080/1547688X.2015.1063178

Chambers, C. (2014, June 10) Physics envy: Do ‘hard’ sciences hold the solution to the

replication crisis in psychology? *The Guardian*. Retrieved from https://www.theguardian.com/science/head-quarters/2014/jun/10/physics-envy-do-hard-sciences-hold-the-solution-to-the-replication-crisis-in-psychology.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.

Crick, N. & Dodge, K. (1999). 'Superiority' is in the eye of the beholder: A comment of Sutton, Smith and Swettenham. *Social Development*, 8, 128-131.

Hawley, P. (2007). Social dominance in childhood and adolescence: Why social competence and aggression may go hand in hand. In P. Hawley, T. Little, & P. Rodkin (Eds). *Aggression and Adaptation: The Bright Side to Bad Behavior* (pp. 1-29). Mahwah, NJ: Erlbaum.

Hawley, P., & Vaughn, B. (2003). Aggression and adaptive functioning: The bright side to bad behavior. *Merrill-Palmer Quarterly*, 49, 239-242.

Ladd, G. (2005). *Children's peer relations and social competence: A century of progress.* New Haven, Yale University Press.

Osborne, J. (2013). Best Practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. Los Angeles, CA: Sage. http://dx.doi.org/10.4135/9781452269948

Osborne, J. (2014). *Best practices in exploratory factor analysis*. Available at: https://dl.dropboxusercontent.com/u/18489687/Osborne_2014_EFA_Proof_revised_protected.pdf

Pepler, D., Craig, W., & Roberts, W. (1995). Social skills training and aggression in the peer group. In J. McCord (Ed.), *Coercion and Punishment in Long-term Perspectives* (pp. 213-228). New York: Cambridge University Press.

Pepler, D., Craig, W., & Roberts, W. (1996, August). *Children's behaviour across contexts: observations of aggressive and nonaggressive children on the playground and in class*. Presented at meetings of the International Society for the Study of Behavioral Development, Quebec City, P.Q, Canada.

Pepler, D., Craig, W., & Roberts, W. (1998). Observations of aggressive and nonaggressive children on the school playground. *Merrill-Palmer Quarterly,* **44**, 55-76.

Roberts, W. (1986). Nonlinear models of development: an example from the socialization of competence. *Child Development*, **57**, 1166-1178.

Roberts, W. (1987). Two-career families: Demographic variables, parenting, and competence in young children. *Canadian Journal of Behavioural Science*, 19, 347-356.

Roberts, W. (2017). False Precision: The Ring of Truth. *Canadian Journal of Behavioural Science*, 49, 97-99. DOI: 10.1037/cbs0000070

Roberts, W. (2020). Parents' observed responses to children's emotional distress: Relations with social competence in preschool. *British Journal of Developmental Psychology*. DOI:10.1111/bjdp.12315

Roberts, W. & Briner, E. (2015, June). Why do parents sometimes become impatient and irritated when their preschool children are emotionally distressed? Presented at meetings of the Canadian Psychological Association, Ottawa, ON.

Roberts, W., & Strayer, J. (1987). Parents' responses to the emotional distress of their children: Relations with children's competence. *Developmental Psychology,* **23**, 415-422.

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development*, 6, 111-135.

Rovenpor, D. & Gonzales, J. (2015). Replicability in psychological science. *Psychological Science Agenda*. https://www.apa.org/science/about/psa/2015/01/replicability.

Rubin, K., Bukowski, W., & Parker, J. (1998). Peer interactions, relationships, and groups. In W. Damon (Ed.), *Handbook of child psychology* (5th ed., Vol. 3, N. Eisenberg, Volume Editor), pp. 619-700. New York, NY: Wiley.

Sousa, D. & Daniel, J. (2018). Qsort R package: A New Tool for Scoring Q-sort Data. *Journal of Open Source Software*, 3(32), 911. https://doi.org/10.21105/joss.00911

Sroufe, L., Egeland, B., Carlson, E., & Collins, W. (2005). *The development of the person: the Minnesota study of risk and adaptation from birth to adulthood*. New York: Guilford Press

Sutton, J., Smith, P., & Swettenham, J. (1999). Socially undesirable need not be incompetent: a response to Crick and Dodge. *Social Development*, 8, 132-134. DOI: 10.1111/1467-9507.00085

Vaughn, B., Santos, A., Monteiro, L., Shin, N., Daniel, J., & Krzysik. (2016). Social Engagement and Adaptive Functioning During Early Childhood: Identifying and Distinguishing Among Subgroups Differing With Regard to Social Engagement. *Developmental Psychology*, 52, 1422-1434. Doi: http://dx.doi.org/10.1037/dev0000142.

Vaughn, B., Shin, N., Kim, M., Coppola, G., Krzysik, L., Santos, A., Peceguina, I., Daniel, J., Veríssimo, M., DeVries, A., Elphick, E., Ballentina, X., Bost, K., Newell, W., Miller, E., Snider, B., & Korth, B. (2009). Hierarchical Models of Social Competence in Preschool Children: A Multisite, Multinational Study. *Child Development*, 80, 1775-1796.

Vaughn, B., Vollenweider, M., Bost, K. & Azria-Evans, M. (2003): Negative interactions and

social competence for preschool children in two samples: Reconsidering the interpretation of aggressive behavior for young children. *Merrill-Palmer Quarterly*, 49, 245-278.

Waters, E., Noyes, D., Vaughn, B. & Ricks, M. (1985). Q-Sort Definitions of Social Competence and Self-Esteem: Discriminant Validity of Related Constructs in Theory and Data. *Developmental Psychology*, 21, 508-522.

Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. *Developmental Review*, 3, 79-97.

Waters, E., Wippman, J., & Sroufe, L. (1979). Attachment, positive affect, and competence in the peer group. *Child Development, 50,* 821-829.

Table 1.

Social Competence: Two Measures with Differing Relations to Social Behaviours

	Social Competence Criterion	Peer Competence	t-test for difference
Roberts (2019), *N*=33			
Peer Competence	0.60	.---	
Bullying	-0.43	0.38	$t(30) = 12.17, p < .0001$
Prosocial (peers)	0.62	-0.10	$t(30) = 8.97, p < .0001$
Cooperative (adults)	0.37	-0.42	$t(30) = 10.42, p < .0001$
Roberts & Strayer (1987), *N*=30			
Peer Competence	0.52	.---	
Bullying	-0.53	0.34	$t(27) = 10.18, p < .0001$
Prosocial (peers)	0.67	0.02	$t(27) = 5.29, p < .0001$
Cooperative (adults)	0.44	-0.29	$t(27) = 5.96, p < .0001$

Note. Significance tests have not been reported for correlations since values replicate, as do the differences described in the text.