Title	Psychometric properties of self-reported quality of life measures for people with intellectual disabilities: A systematic review
Author(s)	Chunxiao Li, Emily W. S. Tsoi, Antony L. Zhang, Shihui Chen and C.K. John Wang
Source	*Journal of Developmental and Physical Disabilities, 25*(2), 253-270
Published by	Springer

Psychometric Properties of Self-reported Quality of Life Measures for People with Intellectual Disabilities: A Systematic Review

Chunxiao Li

Nanyang Technological University

Emily W. S. Tsoi

University of Hong Kong

Antony L. Zhang

Royal Melbourne Institute of Technology

Shihui Chen

Hong Kong Institute of Education

C. K. John Wang

Nanyang Technological University

Author Note

Chunxiao Li and C. K. John Wang are with National Institute of Education, Nanyang Technological University. Emily W. S. Tsoi is with Department of Social Work and Social Administration, University of Hong Kong. Shihui Chen is with Faculty of Arts and Sciences, Hong Kong Institute of Education, Antony L. Zhang is with School of Health Sciences, Royal Melbourne Institute of Technology. Correspodence concerning this manuscript should address Chunxiao Li, Physical Education and Sports Science, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616, Tel: +65 8509 1944, E-mail: cxlilee@gmail.com

Abstract This article aimed to review outcome measure instruments of self-reported Quality of Life (QOL) in people with Intellectual Disability (ID). Instruments were identified from five electronic databases (CINAHL, Medline, PsychINFO, Cochrane Library, and ERIC) and hand-searching with the following inclusion criteria: (i) the paper was to evaluate the measurement properties of QOL; (ii) measure dimensions/indicators that align with the QOL construct; (iii) the target participants should include an ID population; (iv) the measurement should include self report QOL ratings from the ID population; and (v) the study was published in peer-reviewed English language literature. Nine QOL instruments met the inclusion criteria. The evidence of reliability and validity of the included instruments varied. None of the instruments reported for responsiveness. The *Choice Questionnaire* had the most evidence supporting its measurement properties while the *Maryland Ask Me! Project* had the least. Future studies are needed to further refine these instruments.

Psychometric Properties of Self-Reported Quality of Life Measures for People with

Intellectual Disabilities: A Systematic Review

Intellectual disability (ID) is a generalized disorder, which encompasses a wide array of cognitive and behavioral deficiencies; and affects the individuals in all areas of life (AAIDD 2010). The severity of ID is classified into four types on the basis of intelligence quotient test (i.e., mild, moderate, severe, and profound ID; AAIDD 2010). With a childhood onset, ID symptoms persist over the life course. Individuals who have ID are often disadvantaged in equal participation in society and a range of life domains (e.g., health care, family services).

Quality of Life (QOL) is a popular outcome measure in fields such as education, health care, social, and family services (Schalock 2004). Understanding QOL for people with ID is useful for assessing their service needs or the effectiveness of intervention programmes (Schalock 2004). Only when QOL is assessed, areas of lives that need improvement can be found and effective actions (e.g., service design) can then be taken to enhance life quality in ID population (Schalock et al. 2002). Though there are different views with regards to the definition of QOL and its components, there has been a fairly common consensus on the core dimensions and indicators of QOL in ID population (Shalock et al. 2002; Brown et al. 2009). Eight core QOL domains have been identified and cross-culturally validated (Shalock et al. 2002; Brown et al. 2009); and are often mentioned in different QOL instruments. They are social inclusion (e.g., community integration, residential environment), self-determination (e.g., autonomy, choice), personal development (e.g., education and rehabilitation), rights (e.g., privacy, citizenship), interpersonal relations (e.g., friendship, social network), emotional well-being (e.g., contentment, self-concept), physical wellbeing (e.g., health, leisure), and material well-being (e.g., employment, transportation; Cummins 2004b; Shalock et al. 2002; Wang et al. 2010).

These domains are assessed either through objective measures, subjective measures, or a combination of both, depending on the nature of its indicators (Beadle-Brown et al. 2009; Cummins 1997; 2004a). For example, some indicators can be objectively measured through observation and assessed by comparing with general standards (e.g., wealth). Other indicators can only be subjectively evaluated through self-reported experience (e.g., satisfaction). It is believed that subjective QOL cannot be determined from objective measures and there is only a modest relationship between subjective assessment and objective measurement (Cummins 1997; 2005).

Schalock et al. (2002) noted that a valid assessment of QOL must comprise an evaluation of both objective (e.g., direct observation) and subjective (e.g., self-report) components of QOL. However, it is difficult to obtain self-reported QOL ratings from people with ID due to their lower cognitive competence compared with those people without ID (McGillivray et al. 2009). As such, proxy response that involves third parties (e.g., a family member or caretaker) is sometimes utilized to overcome this difficulty (McGillivray et al. 2009). However, most previous research have demonstrated low level of agreement between self-report and proxy-report or modest proxy concordance (Perry and Felce 2002). In addition, some previous studies have repeatedly shown that those with ID generally were reported with lower QOL than those without ID (e.g., Kaptein and Jansen 2008; Watson and Keith 2002). McGillivray et al. (2009), however, found no difference in subjective QOL between people with and without ID. This discrepancy again raises the difficulty in measuring QOL in this population.

Another issue with regards to measuring QOL is the reliability and validity of the self-reported measures. This is of particular concern in a population with ID due to the often present cognitive impairments of the intended subjects. To ask questions requiring a certain level of

cognitive insights and which requires abstract thinking is always difficult for those with impaired intelligence or lower cognitive abilities (Cummins 1997). This has several implications that give rise to the current systematic review. First, the questions set out in the instrument must be kept as simple as possible that minimizes the requirements for abstract understanding (i.e., the face and content validity). Secondly, the fundamental issues of reliability and validity are necessary to address, as pre-testing for the specific population should be required so as to make sure that respondents are able to comprehend the questions.

We witnessed more and more instruments have been developed for measuring QOL in ID population; but the aforementioned issues of proxy report, reliability and validity concerns of QOL measures in such population are often overlooked. While there are studies (e.g., Cummins 1997; Cummins 2004b) that reviewed QOL assessment for persons with ID, narrative reviews were often being used and most of the instruments reviewed in these studies were somewhat out-of-date, all being devised before 2000. Therefore, an up-to-date systematic review is warranted to summarize and appraise the qualities of current self-reported QOL instruments. Thus, the present study aimed to systematically review the existing self-reported QOL measuring instruments targeted for those with ID. Specifically, we intended to answer two research questions as follows: (i) what were the characteristic of these instruments (e.g., number of items, format of response, time to complete, demographic of participants, and sample size)? and (ii) more importantly, were these instruments reliable as well as valid for use in ID population?

Methods

Search Strategy

Articles were identified from a systematic search of five electronic databases from inception until May 2011: CINAHL, Medline, PsychINFO, Cochrane Library, and ERIC. Combinations of the following three groups of keywords were used for the first round searching: (i) Quality of Life; AND (ii) intellectual disability, OR mental retardation, OR Down syndrome, OR learning disabilities; AND (iii) validity, OR reliability, OR psychometric properties. In order to make the search more comprehensive, a second round search was conducted through the name of measurement instruments that were identified in the first round of search using the same databases. Additional articles were also identified through manual checking of the reference lists of included articles and the principal investigator's library and interlibrary link system (i.e., to access a wider range of library materials that cover literature in medicine). Experts in the QOL field were also invited to recommend the extra instruments that were not indentified from the previous two rounds of search.

Inclusion Criteria

Instruments included in this review must satisfy the following criteria: (i) the paper was to evaluate the measurement properties (specify at least one psychometric property) of a QOL instrument; (ii) the instrument should at least measure one of the QOL constructs or indicators that align with at least one of the eight core domains (i.e., social inclusion, self-determination, personal development, rights, interpersonal relations, emotional wellbeing, physical wellbeing, and material wellbeing); (iii) the participants should include an ID population; (iv) the self-reported data (psychometric properties of the QOL measures) specific to ID were reported; and (v) the study was published in peer-reviewed English language literature. Two independent reviewers then excluded irrelevant papers yielding from the initial search by scanning titles and

abstracts. Full text articles were sought if the title and abstract did not provide enough information on whether the inclusion criteria were met. Advice was sought from a third reviewer to reach a consensus when disagreement on paper exclusion/inclusion arose.

Data Extraction

The following data were extracted independently by the two reviewers: name of instrument, dimensions/domains of QOL measurement, number of items, format of response, time to complete, alternate language, country, demographic of participants, severity of ID, reliability, validity, responsiveness, and measurement model (examines if there are problems with floor effects or ceiling effects).

Quality Evaluation

To our best knowledge, an appraisal tool that can be applied specifically to the present review was still limited. The quality of included QOL instruments was assessed by two independent reviewers using the tool that was adapted from the Desirable Characteristics of Outcomes Research Measures for People with Disabilities defined by Andresen (2000). The tool included those recommended assessment criteria for evaluating outcome research questionnaires for disability (e.g., ID, spinal cord injury; Andresen, 2000). Several recent systematic review studies (e.g., Noonan et al. 2009) have used or adapted Andresen's criteria for evaluating the quality of instruments applied in people with disabilities. Nine criteria for measurement properties (three for validity, three for reliability, one for responsiveness, one for instrument bias, and one for measurement model) were used. Table 1 presents a detailed description of the assessment criteria. According to the criteria, grades were assigned to each of the instruments where data

were available. The assigned grades were then summed up to indicate the amount of

psychometric evidence for each QOL measure. Advice from the third reviewer were sought to

reach a consensus whenever agreement over data extraction and quality evaluation cannot be

reached.

Results

Search Results

The initial and subsequent systematic search yielded 309 studies of which 255 were excluded

after reviewing the title and abstract. A further review of fulltexts led to 15 references reporting

nine QOL instruments that met the full inclusion criteria (Fig. 1) and they were: Comprehensive

Quality of Life-Intellectual Disability (ComQol-ID; Cummins 1997), Quality of Life

Questionnaire (QOLQ; Caballo et al. 2005; Rapley and Beyer 1996; Rapley et al. 1994; Schalock

and Keith 1993), Choice Questionnaire (CQ; Stancliffe 1995; Stancliffe and Parmenter 1999),

Health Status Interview Schedule (HSIS; Ruddick and Oliver 2005), Life Satisfaction Scale (LSS;

Heal and Chadsey-Rusch 1985; Yu et al. 1996), Multifaceted Lifestyle Satisfaction Scale (MLSS;

Harner and Heal 1993), Personal Wellbeing Index-Intellectual Disability (PWI-ID; Cummins

and Lau 2005; McGillivray et al. 2009), Maryland Ask Me ! Project (MAMP; Bonham et al.

2004), and Personal Outcome Scale (POS; van Loon et al. 2009).

Measurement Properties

An overview of the participants included in the studies assessing the QOL instruments is

presented in Table 2. The instruments were developed in US (n = 4), Australia (n = 3), UK (n =

1), and Belgium and Holland (n = 1). A total of 2354 adults with mild to profound ID

participated in all of the studies. Instruments that specifically recruited children or adolescents as participants were not found. The range of number of participants varied considerably in the studies, ranging from 21 (HSIS) to 923 (MAMP). Gender distribution was somewhat symmetric, with these studies recruited roughly an equal numbers of males and females as their participants. Most of the instruments (n = 4) included participants with mild to profound ID (QOLQ, MLSS, MAMP, POS), three instruments recruited participants with mild to severe ID (ComQol-ID, CQ, LSS), one instrument was validated using participants with moderate ID (HSIS), and the PWI-ID was developed for participants with mild or moderate ID.

Overview of Instruments

Table 3 presents an overview of the nine included instruments. Three instruments were developed for measuring both subjective and objective QOL (ComQol-ID, QOLQ, POS) in people with ID and the others assessed only the subjective components of QOL. Not all of the included instruments measured the eight core domains of QOL and the number of domains measured by the instruments ranged from four to eight. The number of questions/items in these instruments varied (ranged from 7 in PWI-ID to 58 in MLSS), requiring 20 to 45 mins to complete. All the instruments used rating scale (from 2-point to 11-point) for measuring responses and three measures (ComQol-ID, PWI-ID, MAMP) applied graphic rating scales for helping participants to understand questions.

Overview of Reliability and Validity

Table 4 shows the reliability evidence regarding the instruments. Six of the instruments (QOLQ, CQ, LSS, MLSS, ComQol-ID, PWI-ID) reported the overall consistency reliability. Overall test-

retest reliability has been reported for five instruments (ComQol-ID, QOLQ, LSS, MLSS, PWI-ID). Evidence of inter-rater (staff or interviewers) reliability was found in four instruments (QOLQ, LSS, MLSS, POS) but the POS did not report the overall value. Two instruments (PWI-ID, MAMP) although were administered through interview (Table 4), did not report the inter-rater reliability. Only three instruments (ComQol-ID, CQ, POS) reported correlation coefficients between consumer and proxy report. Other instruments (QOLQ, HSIS, LSS, MLSS, PWI-ID, MAMP) with target participants of moderate to profound ID, did not provide any self-proxy data.

The evidence for validity of the instruments is summarized in Table 5. Only two instruments (QOLQ, PWI-ID) provided the evidence of their factor structure through exploratory factor analysis. The convergent validity was available in four instruments (QOLQ, CQ, MLSS, POS). All the instruments reported content validity. Three instruments (CQ, MAMP, POS) involved people with ID for initial/pilot testing. Most of the instruments (n = 7) were able to provide evidence with regards to discriminant validity. Floor and ceiling effects was only observed in the CQ (i.e., able to reduce the occurrence of both ceiling and floor effects), which means the other instruments (n = 8) did not report this critical psychometric data.

Further Description on Instruments

The ComQol-ID measured both subjective and objective QOL tapping seven domains. The instrument was translated to most major languages (e.g., Spanish and Chinese). One feature of this instrument was the use of a pretesting protocol to help practitioners select appropriate graphic rating scales (2-, 3-, or 5- point) for participants. In terms of reliability, internal consistency of the ComQol-ID was weak (overall α = .56 for subjective measures, overall α = .68 for objective measures). Though overall test-retest reliability within a 2-week period was strong

(overall *r* = .87 for subjective measures, overall *r* = .82 for objective measures), some of its subscales indicated low test-retest reliabilities (*r* = .12-.97). Self- and proxy-report correlation was only found in three of its domains and the relationships were weak (health, *r* = .21; productivity, *r* = .08; safety, *r* = -.02). With regards to content validity, the ComQol-ID was developed in terms of the current definition of QOL (enough evidence on content validity). Factorial and convergent validity of the instrument were not examined.

Similar with the ComQol-ID, the QOLQ also measured both subjective and objective QOL but on four domains through 3-point rating scales. Other than English language version, the instrument was also available in Spanish and Mexican languages. The QOLQ had acceptable to excellent internal reliability (*r* = .67-.90) and inter-rater reliability (*r* = .66-.83). Excellent test-retest reliability within a 2-week interval was found (*r* = .80-.96). However, the proxy concordance information was not reported. The QOLQ provided plenty of information on validity. The factorial validity was examined through exploratory factor analysis; however, the four domains only explained 33.7% of the variance. The content validity of the QOLQ was deemed enough and it moderately correlated with the MLSS (*r* = .57). Its dscriminant validity was excellent.

The CQ measured only subjective QOL domains (n = 6) through 3-point rating scales. The CQ recruited only 30 participants for the validation study, limiting the report of information on its psychometric properties. The CQ demonstrated excellent overall internal reliability (*r* = .81) and had overall moderate concordance between consumer and proxy report (*r* = .58 for staff 1; *r* = .49 for staff 2). The CQ demonstrated excellent convergent correlations with QOLQ (*r* = .79-.83) and acceptable content validity. Its dscriminant validity was deemed excellent. Information of its test-retest reliability and factorial validity were not available.

Eight subjective domains were measured in the HSIS. Similar to the CQ, small sample size (n = 21) was used in its validation study. The HSIS had low to high internal reliability across its domains (*r* = .43-.82) and low to moderate test-retest reliability (ICC = .22-.66). The instrument showed excellent content validity. However, the HSIS failed to provide information on proxy concordance, factorial validity, convergent validity, and discrminant validity.

The LSS used only 38 participants in its validation study. This instrument measured four subjective QOL domains through Yes/No responses plus an open-ended format. The response format may carry measuring problems (e.g., bias on positive rating). It had acceptable to excellent internal consistency across the four domains (*r* = .64-.85) and excellent overall internal reliability (*r* = .81). The instrument had moderate to strong test-retest reliability within a period of 4-29 days (*r* = .44-83). Excellent inter-rater reliability was reported (ICC = .82-.99). However, information on self- and proxy- report correlation was not provided. Acceptable content validity and excellent dscriminant validity were found in the LSS. Evidence on factorial validity and convergent validity were not reported.

Deriving from the LSS, the MLSS also measured subjective QOL but on seven domains using 5-point rating scales. Similar to the LSS, the MLSS showed low to excellent internal consistency across the four domains (*r* = .56-.92) and excellent overall internal reliability (*r* = .88). Although the LSS had excellent overall test-retest reliability within a period of 6-64 days, it was found that some of its domains had low reliability (*r* = .11-.85). Excellent inter-rater reliability was found in the LSS (ICC = .91-.97). The convergent validity was examined in the MLSS; however, it showed no correlation with QOLQ. Acceptable content and excellent dscriminant validity were found in the MLSS. The instrument did not report evidence on proxy concordance and factorial reliability.

The PWI-ID assessed seven subjective QOL domains. The PWI-ID was another instrument integrating pretesting protocol to determine the use of types of graphic rating scales (2-, 3-, 5-, or 11- point). It seems that using the 11-point rating scale could be too long for participants with ID. The PWI-ID had moderate internal consistency ($r = .76$) and moderate test-retest reliability within a two-week interval (ICC = .57). Its factorial validity was evident through an exploratory factor analysis, explaining about 57.97% of the variance. Moderate discrminant and content validity were found in the PWI-ID. Information on self- and proxy-report reliability and convergent validity were not offered.

The MAMP measured eight subjective QOL domains through graphic 3-point rating scales. A very positive side of the MAMP was that it recruited a large sample size (n = 923) for its validation study. Despite of the large sample size, factorial validity was not reported, nor weretest-retest reliability, proxy concordance, convergent validity, and discrminant validity. The MAMP had moderate internal reliability across its domains ($r = .70-.76$) and excellent content validity.

The POS was the most recently developed instrument. The instrument measured both subjective and objective QOL domains (n = 8) through 3-point rating scales. In terms of reliability, the POS demonstrated a moderate to excellent concordance between consumer and proxy report ($r = .42-.80$) and it had mild to moderate internal reliability ($r = .25-.75$) and inter-rater reliability ($r = .00-.79$) across its domains. Information on test-retest reliability information was lacking. There was mild to moderate correlation between the POS and the GENCAT-Scale (Verdugo et al. 2010). The POS showed excellent content and discriminant validity, but the factorial validity evidence was not reported.

Quality Evaluation

Table 6 presents summary of grades assigned for each of the instruments. Inter-rater reliability of instrument rating between two independent reviewers was .87. The responsiveness data of the included instruments was not reported as such information was not available. The CQ and QOLQ had relatively more evidence supporting its measurement properties. In addition, CQ and QOLQ were assigned for five (4A and 2B) and six (2A, 3B, 1A-C) of the eight measurement properties respectively. On the other hand, MAMP had the least amount of evidence with only two measurement properties (1A, 1B).

Discussion

An ideal instrument assessing QOL in people with ID should be one which is valid, reliable, with low administer burden, and cultural-specific to the target population involved. As such, this study systematically reviewed the related instruments. Our study found nine instruments assessing a variety of QOL constructs for this population. More importantly, there were varying amount of information provided by the authors regarding the psychometric properties of the included instruments, with the CQ providing the most information while the MAMP offering the least information.. Regardless of the psychometric evidence, eight instruments were validated in English speaking populations (Australia, UK, and US) and one (Belgium and Holland) used non-English speaking participants in the validation procedure. Even though ComQol-ID and QOLQ were the only ones validated and translated in other major languages (e.g., Spanish and Chinese), other instruments cover domains that these two instruments do not due to differing measuring purposes. Consequently, cross-culture validations of all instruments are recommended.

Evidence showing internal consistency was available in all the instruments but only four out of them (QOLQ, CQ, LSS, MLSS) had an excellent overall rating. This piece of information should be interpreted with caution as these scales contained varying number of items (e.g., a scale with more items tends to have a better α level) and they also used participants with varying characteristics (e.g., some studies recruited more people with severe or profound ID compared with others that can obviously impact the internal reliability). There were also insufficient investigations in factorial structures of the instruments with only two performed exploratory factor analysis. None of the instruments applied confirmatory factor analysis or more advanced measurement methods (e.g., Rasch analysis which is based on Item Response Theory; Fischer and Molenaar 1995) for examining factorial structure (Noonan et al. 2009).

Subjective QOL is an important but challenging aspect to measure in people with ID (Cummins 2004a; McGillivray et al. 2009). In terms of the operation of the concept of QOL, all the instruments were developed to measure subjective QOL with three of them (ComQol-ID, QOLQ, POS) were also able to measure the objective components. All of the instruments used adults or older adults with ID to for testing of psychometric properties, their uses in children or adolescents with ID are still unknown. Moreover, regarding the characteristics of participants, the HSIS was the only instrument used specific severity level of ID (moderate) for validation.

Evidence regarding the proxy concordance was only available in three instruments. It could be because of the argument on the validity of the proxy response especially for those individuals with severe to profound ID (Beadle-Brown et al. 2009; Cummins 1998). Response bias could be due to this group of people suffered difficulties such as production of symbols and construction of concepts (Finlay and Lyons 2001). That (low cognitive competence) could be in particular the case for measuring subjective QOL. Cummins (1998) found that it is better to use

proxy-report to evaluate objective QOL (e.g., wealth) rather than subjective QOL (e.g., satisfaction) as it may be unreliable to use proxy-report to assess subjective QOL. In addition, despite potential responding bias in people with ID; interviewer/staff's bias (e.g., overestimation, unfamiliar with consumers) should not be overlooked (Stancliffe 1999). In these cases, a Rasch analysis would be helpful to detect the bias (e.g., person fit statistics).

Even with the potential for bias by the proxies, still many scholars have recommended involving proxies who know the consumer well for QOL assessment (e.g., Cummins 1997; Schwartz and Rabinovitz 2003); however, no screening tool was used in related studies to check to what extent they know the consumer in order to exclude the "noise" that may impact on calculating proxy-self agreement. Regardless of the possible low correlation between proxy and self-reported ratings, Stancliffe (1999) suggested that proxy responses should still be used. Schwartz and Rabinovitz (2003) added that including proxy data is still of use for generating a more holistic picture of the QOL of people with ID. In fact, proxy concordance could be better achieved through using two independent proxy respondents than one (Bonham et al. 2004; Schalock and Keith 1993; Stancliffe 1999).

A consideration of the time required to complete the questionnaire is an important issue for measurement in people with disabilities. It was suggested by Andresen (2000) that measures designed for people with disabilities should be brief and ideally be less than 15 min. However, none of those instruments with available data on required completion time can satisfy Andresen's recommendation. This could be due to some of the instruments (ComQol-ID, PWI-ID) contained structured pretesting protocol for determining participants' cognitive performance and thus choosing the appropriate type of rating scale, which may substantially lengthen the time required to complete. Nevertheless, the inclusion of a pretesting protocol even though will surely

increase administration time of the instrument, it is still essential as it could improve validity of the scale (Petry and Maes 2006). An alternative way to decrease administration time of the instrument while maximizing validity could be to use graphical rating scale (Hartley and Maclean 2006). Although various rating scales (from 2- to 11-point) were used in all of the instruments, only two instruments (ComQol-ID, PWI-ID) adopted graphical rating scales.

Content validity/ item bias were assessed in all the instruments, but only three (HSIS, MAMP, POS) of them involved persons with ID in developing the content or their initial items. Pirpiris and Graham (2004) noted that the unique characteristics of population (i.e., people with ID) should be considered in the development of instruments. Unfortunately, it seems that most of the identified instruments in current study were not able to address this issue. Moreover, a lack of good content or face validity may also impact negatively on the test-retest reliability of the instruments.

Only four (QOLQ, CQ, MLSS, POS) out of nine instruments tested their convergent validity. This could be due to a lack of reliable and valid instruments on measuring QOL in people with ID for use to correlate with their own instruments (McGillivray et al. 2009). The QOLQ were used in two studies for examining convergent correlations since the QOLQ was one of the more credible instruments compared with other QOL measure for ID population (Cummins, 1997). It is perplexing as to why the QOLQ correlated with the MLSS (Schalock and Keith 1993), but not vice versa (Harner and Heal 1993; Table 5). This may be because of the validity of these two QOL questionnaires had not been thoroughly studied. Another explanation could be the varying sample size (n = 552 [with correlation] versus n = 149 [without correlation]), which implies that the psychometric properties can vary a lot, even when we examine the same instruments.

Study Limitations

The present study has two major limitations. Firstly, although this systematic review was able to summarize evidence for the methodological quality of the reviewed self-reported QOL measures, we are not able to evaluate all of the desirable characteristics of an ideal QOL measurement (e.g., administrator burden; Andresen 2000). Secondly, the exclusion of non English language articles may preclude some potential relevant instruments to be included in the current review.

Implications

Despite of the limitations, several implications could be drawn from this study. Firstly, this study provides an overview and evidence for guiding researchers and practitioners to choose an appropriate self-reported QOL measure for use in ID population. Choosing a reliable and validate instrument is of importance to assess intervention effectiveness for people with ID. Secondly, although some instruments (e.g., CQ and QOLQ), at least in terms of our present findings, show promising in measuring self-reported QOL in people with ID, further refinement is still much needed for most of the instruments, in particular the need for documenting the responsiveness of the included instruments as none of the instruments provided the evidence on responsiveness to treatment, which therapists or researchers may use for assessing if their interventions are effective. Thirdly, we also recommend modern measurement theory model such as the Item Reponse Theory (IRT) be taken into account in their development (Fischer and Molenaar 1995). For example, using IRT to examine whether there are floor and ceiling effects (the CQ was the only instrument reported the floor and ceiling effect in current study). Lastly, instruments should be validated by involving participants with differing level of severity of ID.

In addition, pre-testing protocol could be integrated to QOL instruments to determine participants' cognitive performance, as they can be used to guide staff or interviewers to choose an appropriate type of rating scales (e.g., 3-point rating scale vs. 5-point rating scale) and thus improving the validity or reliability of the measuring.

In summary, there is a wealth of measuring instruments on QOL in ID population. Unfortunately, most of the instruments were not well validated. Taking into account the psychometric evidence reported in this review, the CQ and QOLQ were found to be the relatively more superior scales. This conclusion should be taken cautiously as the main purpose of the present review was not to 'compare' QOL instruments and it was difficult to compare them because of the heterogeneity of the included measures (different domains of QOL, rating scales, sample sizes etc.); rather, the current review was presented hoping to shed some insights for practitioners to select an appropriate instrument and to pinpoint directions for researchers to further develop or refine the instruments in this field.

References

AAIDD (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th
   ed). Retrieved from http://www.aaidd.org/media/PDFs/CoreSlide.pdf

Aman, M.G., & Singh, N. N. (1986). *Aberrant behavior checklist manual*. New York: Slosson
   Educational Publications.

Andresen, E. M. (2000). Criteria for assessing the tools of disability outcomes research. *Archives
   of Physical Medicine and Rehabilitation*, *81*, S15-S20.

Bonham, G. S., Basehart, S., Schalock, R. L., Marchand, C.B., Kirchner, N., Rumenap, J. M., &
   Scotti, J. (2004). Consumer-based quality of life assessment: The Maryland Ask Me!
   Project. *Mental Retardation*, *42*, 338-355.

Beadle-Brown, J., Murphy, G., & DiTerlizzi, M. (2009). Quality of Life for the
   Camberwell Cohort. *Journal of Applied Research in Intellectual Disabilities*, *22*,
   380-390.

Brown, R. I., Schalock, R. L., & Brown, I. (2009). Quality of life: Its application to persons with
   intellectual disabilities and their families- introduction and overview. *Journal of Policy and
   Practices in Intellectual Disability*, *6*, 2-6.

Caballo, C., Crespo, M., Jenaro, C., Verdugo, M. A., & Martinez, J. L. (2005). Factor structure
   of the Schalock and Keith Quality of Life Structure (QOL-Q): Validation on Mexican and
   Spanish samples. *Journal of Intellectual Disability Research*, *49*, 773-776.

Cummins, R. A. (1997). Self-rated Quality of Life scales for people with an intellectual
   disability: a review. *Journal of Applied Research in Intellectual Disabilities*, *10*, 199-
   216.

Cummins R. A. (1998). *Measuring subjective life quality by proxy response*. Paper

presented at the Australian Society for the Study of Intellectual Disability National Conference, Adelaide, South Australia, 28 September 1998.

Cummins, R. A. (2004a). Self-rated quality of life scales for people with an intellectual disability: A reply to Ager & Hatton. *Journal of Applied Research in Intellectual Disabilities*, *14*, 1-11.

Cummins, R. A. (2004b). Instruments assessing quality of life: Characteristics and functions. In J. H. Hogg & A. Langa (Eds.), *Approaches to the assessment of adults with intellectual disabilities: Part I: A service provider's guide* (pp. 130-152). London: Blackwell Publishing.

Cummins, R. A., & Lau, A. D. L. (2005). *Personal Wellbeing Index-Intellectual Disability* (3rd ed). Australia: Deakin University.

Finlay, W. M. L., & Lyons, E. (2001). Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. *Psychological Assessment*, *13*, 319-335.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch Models: Foundations, recent developments, and applications*. New York: Springer.

Harner, C. J., & Heal, L. W. (1993). The Multifaceted Lifestyle Satisfaction Scale (MLSS): psychometric properties of an interview schedule for assessing personal satisfaction of adults with limited intelligence. *Research in Developmental Disabilities*, *14*, 221-236.

Hartley, S. L., & Maclean, W. E. (2006). A review of the reliability and validity of Likert-type scales for people with intellectual disability. *Journal of Intellectual Disability*, *50*, 813-827.

Heal, L.W., & Chadsey-Rusch, J. (1985). The Lifestyle Satisfaction Scale (LSS): Assessing individuals' satisfaction with residence, community settings an associated services. *Applied*

Research in Mental Retardation, *6*, 470-495.

Kaptein, S., & Jansen, D. (2008). Mental health problems in children with intellectual disability: use of the strengths and difficulties questionnaire. *Journal of Intellectual Disability Research*, *52*, 125-131.

McGillivray, J. A., Lau, A. L. D., Cummins, R. A., & Davey, G. (2009). The utility of the Personal Wellbeing Index Intellectual Disability Scale in an Australian Sample. *Journal of Applied Research in Intellectual Disabilities*, *22*, 276-286.

Noonan, V. K., Miller, W. C., & Noreau, L. (2009). A review of instruments assessing participation in persons with spinal cord injury. *Spinal Cord*, *47*, 435-446.

Nota, L., & Soresi, S. (2002). L'analisi della Qualita della Vita di persone adulte con ritardo mentale: proposta di uno strumento di eterovalutazione [The quality of life of adults with mental retardation]. *Giornale Italiano Delle Disabilta [Italian Journal of Disability]*, *2*, 10-26.

Nota, L., Soresi, S., & Perry, J. (2006). Quality of life in adults with an intellectual disability: The Evaluation of Quality of Life Instrument. *Journal of Intellectual Disability Research*, *50*, 371-385.

Novak, A. R., Heal, L. W., Pilewski, M. B., Laidlaw, T. (1980). *Independent apartment settings for developmentally disabled adults: An empirical analysis*. Paper presented at the meeting of the American Association on Mental Deficiency, San Francisco.

Ouellette-Kuntz, H. (1990). A pilot study in the use of the Quality of Life Interview Schedule. *Social Indicators Research*, *23*, 283-298.

Ouellette-Kuntz, H., & McCreary, B. D. (1996). *Quality of life assessment for persons with severe developmental disabilities*. In R. Renwick, I. Brown and M. Nagler (eds) Quality of

life in health promotion and rehabilitation: Conceptual approaches, issues and applications. Thousand Oaks, CA: Sage.

Ouellette-Kuntz, H., McCreary, B. B, Minnes, P., & Stanton, B. (1994). Evaluating quality of life: The development of the Quality of Life Interview Schedule (QUOLIS). *Journal on Developmental Disabilities*, *3*, 17-31.

People on the Go. (1996). *Signs of quality*. Annapolis: The Arc of Maryland.

Petry, K., Kuppens, S., Vos, P., & Maes, B. (2010). Psychometric evaluation of the Dutch version of the Mood, Interest and Pleasure Questionnaire (MIPQ). *Research in Developmental Disabilities*, *31*, 1652-1658.

Petry, K., & Maes B. (2006). Identifying expressions of pleasure and displeasure by persons with profound multiple disabilities. *Journal of Intellectual and Developmental Disability*, *31*, 28-38.

Perkins, A. (2007). Self- and proxy reports across three populations: Older adults, Persons with Alzheimer's Disease, and persons with intellectual disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, *4*, 1-10.

Pirpiris, M., & Graham, H. K. (2004). Uptime in children with cerebral palsy. *Journal of Pediatric Orthopaedics*, *24*, 521-528.

Rapley, M., & Beyer, S. (1996). Daily activity, community participation and quality of life in an ordinary housing network. *Journal of Applied Research in Intellectual Disabilities*, *9*, 31-39.

Rapley, M., Lobley, J., & Bozatzis, N. (1994). *Preliminary validation of the Schalock and Keith (1994) Quality of Life Questionnaire with a British Population*. Lancaster: Department of Psychology, Lancaster University.

Ross, E., & Oliver, C. (2003). Preliminary analysis of the psychometric properties of the Mood, Interest & Pleasure Questionnaire (MIPQ) for adults with severe and profound learning disabilities. *British Journal of Clinical Psychology*, *42*, 81-93.

Ruddick, L., & Oliver, C. (2005). The development of a health status measure for self-report by people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, *18*, 143-150.

Schalock, R. L. (2004). The concept of quality of life: what we know and not know. *Journal of Intellectual Disability Research*, *48*, 203-216.

Schalock, R. L., Brown, I., Brown, R., Cummins, R. A., Felce, D., Matikka, L., et al. (2002). Conceptualization, measurement, and application of quality of life for persons with intellectual disabilities: Report of an international panel of experts. *Mental Retardation*, *40*, 457-470.

Schalock, R.L., & Keith, K.D. (1993). *Quality of Life Questionnaire Manual*. Worthington, Ohio: IDS Publishing Corporation.

Schalock, R. L., Keith, K. D., Hoffman, K., & Karan, O. C. (1989). Quality of Life: Its measurement and use. *Mental Retardation*, *27*, 25-31.

Schwartz, C., & Rabinovitz, S. (2003). Life satisfaction of people with intellectual disability living in community residences: perceptions of the residents, their parents and staff members. *Journal of Intellectual Disability Research*, *47*, 75-84.

Stancliffe, R. J. (1995). *Choice and decision making and adults with intellectual disability*. Unpublished doctoral thesis, Macquarie University, Sydney.

Stancliffe, R.J., & Parmenter, T.R. (1999). The Choice Questionnaire: A scale to assess choices exercised by adults with intellectual disability. *Journal of Intellectual & Developmental*

Disability, *24*, 107-132.

van Loon, J., van Hove, G., Schalock, R., & Claes, C. (2009). *Personal Outcome Scale: Administration and standardization manual*. Retrieved on 16 Jan 2012, from http://www.poswebsite.org/docs/POS-Admin_-Stand__Manual.pdf

Verdugo, M. Á., Arias, B., Gómez, L. E., & Schalock, R. L. (2010). Development of an objective instrument to assess quality of life in social services: Reliability and validity in Spain. *International Journal of Clinical and Health Psychology*, *10*, 105-123.

Wang, M., Schalock, R. L., Verdugo, M. A., & Jenaro, C. (2010). Examining the factor structure and hierarchical nature of the quality of life construct. *American Journal on Intellectual and Developmental Disabilities*, *115*, 218-233.

Watson, S. M. R., & Keith, K. D. (2002). Comparing the quality of life of school-age children with and without Disabilities. *Mental Retardation*, *40*, 304-312.

Ware, J.E., Snow, K.K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey. Manual and Interpretation Guide*. The Health Institute, New England Medical Centre, Boston.

Yu, A. L. C., Jupp, J. J., & Taylor, A. (1996). The discriminate validity of the Lifestyle Satisfaction Scale (LSS) for the assessment of Australian adults with intellectual disabilities. *Journal of Intellectual and Developmental Disability*, *21*, 3-15.

Table 1

Criteria for Instrument Assessment

Criteria	Description	Evaluation
Reliability	The degree to which an instrument is consistent or free from random error	Internal consistency (coefficient α) A$\geq$.80; B>.70, <.80; C$\leq$.70
		Test-retest repeatability (ICC/κ, *r*) A$\geq$.75; B >.40, < .75; C$\leq$.40
		Proxy responses (ICC/κ, *r*) A$\geq$.75; B >.40, <.75; C$\leq$.40
Validity	The degree to which an instrument measures what it intends to measure. Types include factorial structure (assess dimensionality); convergent correlations (comparisons with other measures); and discriminant (differentiate based on known groups)	Factorial structure (exploratory or confirmatory factor analysis, Rasch analysis) A=confirmed, Rasch analysis is good B=factorial analysis is good or Rasch has some problems C=inadequate statistical analysis
		Convergent correlations A$\geq$.60; B>.30, <.60; C$\leq$.30
		Discriminant (differences by means or %) A=strong, in expected direction B=moderate or conflicting evidence C=weak

Table 1

Criteria for Instrument Assessment (*Continued*)

Responsiveness	The ability of an instrument to measure important changes following intervention(s)	A=strong, in expected direction B=moderate or conflicting evidence C=weak or based only on statistical methods
Item bias	Assesses in practical terms if individual questions or summary scores are biased for individuals with ID	A=persons with ID reviewed the instrument and acceptability is published B=there is adequate face validity to support low bias C=bias is evident
Measurement model	Examines if there are problems with floor effects (lowest level of ability) or ceiling effects (highest level of ability).	The instrument has scales or measures where 20% of persons with ID are grouped at scoring extremes. Also, can consider the score distribution (mean and standard deviation). A=no problems B=few or marginal problems C=substantial skewing of scales/measures

Note. ICC=intraclass correlation coefficient; α=coefficient alpha; κ=kappa; r= Pearson's r.

Adapted from "Criteria for assessing the tools of disability outcomes research", by E. M. Andresen, 2000, *Archives of Physical Medicine and Rehabilitation*, *81*, 16.

Table 2

Description of Study Participants

Study	Country	NO of Participants	Age (Year)	Gender	Type of ID
ComQol-ID (Cummins 1997)	Australia	59 69	M=37.2,SD=13.5 M=21.4,SD=6.8	Gender balance F=74%	Mild to severe
QOLQ (Schalock and Keith 1993)	US	552 (Nebraska I =312, Nebraska II =145, Colorado=95)	Nebraska I =38.5 Nebraska II=38.9 Colorado=36.1	F=44.1% F=49.0% F=47.4%	Mild to profound
CQ (Stancliffe and Parmenter 1999)	Australia	30	M=35.6,SD=12.1	M=15 F=15	Mild to severe
HSIS (Ruddick and Oliver 2005)	UK	21	M=46.7, SD=7.0	M=8 F=13	Moderate
LSS (Heal and Chadsey-Rusch 1985)	US	38	Ranged from 20-67	M=19 F=19	Mild to severe

Table 2

Description of Study Participants (Continued)

MLSS (Harner and Heal 1993)	US	149	M=40	M=58% F=42%	Mild to profound
PWI-ID (Cummins and Lau 2005)	Australia	114	M=33	M=62 (54%) F=52 (46%)	Mild to moderate
MAMP (Bonham et al. 2004)	US	923	Ranged from 18-64 (90%), ≥ 65 (6%), Missing (4%)	ND	Mild to profound
POS (van Loon et al. 2009)	Belgium and Holland	399	ND	ND	Mild to profound

Note. ComQol-ID=Comprehensive Quality of Life-Intellectual Disability; QOLQ=Quality of Life Questionnaire; CQ=Choice Questionnaire; HSIS=Health Status Interview Schedule; LSS=Life Satisfaction Scale; MLSS=Multifaceted Lifestyle Satisfaction Scale; PWI-ID=Personal Wellbeing Index-Intellectual Disability; Maryland Ask Me! Project = MAMP; and POS=Personal Outcomes Scale. ID=Intellectual Disability; ND=No Data

Table 3

Description of Participants Instruments

Instrument	QOL domains, Subjective or objective QOL	NO of items, Rating scale, Time to complete
ComQol-ID (Cummins 1997)	QOL domains (n=7): material wellbeing; health; productivity; intimacy; safety; community; and emotional well-being Subjective/objective: both	Item: 35 + pretesting protocol Rating:2, 3, or 5 point graphic rating scales base on the performance in pretesting protocol Time: 45 min
QOLQ (Schalock and Keith 1993)	QOL domains (n=4): empowerment/independent; competency/productivity; satisfaction; and social belonging/community integration Subjective/objective: both	Item: 40 Rating: 3-point scoring system Time: 20-30 min
CQ (Stancliffe and Parmenter 1999)	QOL domains (n=6): domestic activities; co-residents and staff; money and spending; health; social activities, community access and personal relationships; work and day activities; overall choice Subjective/objective: subjective	Item: 26 Rating: 3-point scale Time: ND

Table 3

Description of Participants Instruments (Continued)

HSIS (Ruddick and Oliver 2005)	QOL domains (n=8): physical function; bodily pain; general health; vitality; social functioning; mental health; reported health transition; sensory functioning; and memory functioning Subjective/objective: subjective	Item: 26 Rating: rating scale, forced-choice items, open comments Time: ND
LSS (Heal and Chadsey-Rusch 1985)	QOL domains (n=4): community satisfaction; friends and free-time satisfaction; satisfaction with services; general satisfaction; and job satisfaction Subjective/objective: subjective	Item: 29 Rating: Yes-No questions Time: 20 min
MLSS (Harner and Heal 1993)	QOL domains (n=7): community satisfaction; friends and free time satisfaction; job satisfaction; recreation and leisure satisfaction ; client control and self-determination; and general satisfaction Subjective/objective: subjective	Item: 58 Rating: 5-point rating scale Time: 20 min

Table 3

Description of Participants Instruments (Continued)

PWI-ID (Cummins and Lau 2005)	QOL domains (n=7): standard of living; personal health; achievement in life; personal relationship; personal safety; community-connectedness; future security; and religion/spirituality Subjective/objective: subjective	Item: 7, pretesting protocol Rating: graphic rating scales (2-, 3-, 5-, or 11-point) depending on respondents' cognitive level Time: 45min
MAMP (Bonham et al. 2004)	QOL domains (n=8):social inclusion, self-determination, personal development, rights, interpersonal relations, emotional wellbeing, physical wellbeing, material wellbeing Subjective/objective: subjective	Item: 56 Rating: favorable, neutral, unfavorable (graphic) Time: 20-30 min
POS (van Loon et al. 2009)	QOL domains (n=8): social inclusion, self-determination, personal development, rights, interpersonal relations, emotional wellbeing, physical wellbeing, material wellbeing Subjective/objective: both	Item: 48 Rating: 3-point rating scale Time: ND

Note. ND=No Data

Table 4

Reliability of the Participation Instruments

Instrument	Internal consistency (α)	Test-retest period and Coefficients	Inter-rater time period and Coefficients
ComQol-ID (Cummins 1997)	.11-.65 across all domains; overall .56 for subjective scale and .68 for objective scale	Period: 1-2 weeks Coefficients: r=.12-.97 across all domains; overall r=.87 for subjective scale and r=.82 for objective scale	Time: ND Coefficients: ND
QOLQ (Schalock and Keith 1993)	.67-.90 across all domains; overall .90	Period: 2 weeks Coefficients: r=.80 to .96 across all domains; overall r=.87	Time: two independent raters rated the questionnaire Coefficients: r=.66-.83 across all domains; overall r=.83
CQ (Stancliffe and Parmenter 1999)	Overall .81	Period: ND Coefficients: ND	Time: ND Coefficients: ND

Table 4

Reliability of the Participation Instruments (Continued)

HSIS (Ruddick and Oliver 2005)	.43-.82 across all domains	Period: ND Coefficients: ICC=.22-66 across all domains	Time: ND Coefficients: ND
LSS (Heal and Chadsey-Rusch 1985)	.64-.85 across all domains; overall .81	Period: 4-29 days Coefficients: r=.44-83 across all domains; overall r=.74	Time: ND Coefficients: ICC=.82-.99 across all domains; overall ICC=.95
MLSS (Harner and Heal 1993)	.56-.92 across all domains; overall .88	Period: 6-64 days Coefficients: r=.11-85 across all domains; overall r=.70	Time: two raters scoring the same on-site interview Coefficients: ICC=.91-.97 across all domains; overall ICC=.99
PWI-ID (Cummins and Lau 2005)	Overall .76	Period: 1-2 weeks Coefficients: overall ICC=.57, r=.58	Time: ND Coefficients: ND

Table 4

Reliability of the Participation Instruments (Continued)

MAMP (Bonham et al. 2004)	.70-.76 across all domains	Period: ND Coefficients: ND	Time: ND Coefficients: ND
POS (van Loon et al. 2009)	.25-.75 across all domains	Period: ND Coefficients: ND	Time: ND Coefficients: r=.67-.79 across all domains except Self-determination domain found not relationship

Note. ND=No Data

Table 5

Validity and Related Measurement Properties of the Participation Instruments

Instrument	Validity: factorial structure, convergent correlations, discriminant group	Item bias (content validity)
ComQol-ID (Cummins 1997)	Factorial: ND Convergent: ND Discriminant: discriminate college subjects without ID	The questionnaire incorporates a contemporary understanding of QOL construct
QOLQ (Schalock and Keith 1993)	Factorial: four factors explained 33.7% of total variance through EFA Convergent: a correlation of .57 with the MLSS(Harner and Heal 1993) Discriminant: discriminate IQ, living and work environment	Items were generated published sources on well-being, independence)
CQ (Stancliffe and Parmenter 1999)	Factorial: ND Convergent: correlations with QOLQ (Schalock and Keith 1993) were .79 Discriminant: discriminate participants lived either semi-independents or in group homes	items were drawn from comprehensive review and following expert review and empirical evaluation

Table 5

Validity and Related Measurement Properties of the Participation Instruments (Continued)

CQ (Stancliffe and Parmenter 1999)	Factorial: ND Convergent: correlated with QOLQ (*r* = .79; Schalock and Keith 1993) Discriminant: discriminate participants lived either semi-independents or in group homes	items were drawn from comprehensive review and following expert review and empirical evaluation
HSIS (Ruddick and Oliver 2005)	Factorial: ND Convergent: ND Discriminant: ND	Developed by adapting questions items and response formats from SF-36 (Ware et al. 1993); a pilot was conducted
LSS (Heal and Chadsey-Rusch 1985)	Factorial: ND Convergent: ND Discriminant: discriminate living environments (i.e., care facilities and apartments, community living)	The LSS was refined from Residential Satisfaction Scale (Novak et al. 1980)
MLSS (Harner and Heal 1993)	Factorial: ND Convergent: Not correlated with QOLQ (Schalock and Keith 1993) Discriminant: discriminate age, level of ID & residential restrictiveness	MLSS was modified from LSS which was refined from Residential Satisfaction Scale (Novak et al. 1980)

Table 5

Validity and Related Measurement Properties of the Participation Instruments (Continued)

PWI-ID (Cummins and Lau 2005)	Factorial: two factors explained about 57.97% of the variance through EFA Convergent: ND Discriminant: discriminate different age groups	Collaborative and international network of scholars to develop and use it
MAMP (Bonham et al. 2004)	Factorial: ND Convergent: ND Discriminant: ND	3/4 of the questions came from a list of questions generated by people receiving services ("People on the Go" 1996); pilots
POS (van Loon et al. 2009)	Factorial: ND Convergent: seven of the domains were moderately correlated with GENCAT-Scale (Verdugo et al. 2010) except the material wellbeing domain Discriminant: there is a significant difference among the four levels of intellectual functioning	Items were derived from the literature, expert review, and focus groups; pilots were conducted

Note. EFA=Exploratory Factor Analysis; ND=No Data

Table 6

Summary of the Measurement Properties

Instrument	Internal consistency	Test-retest reliability	Proxy concordance	Factorial validity	Convergent validity	Discriminant validity	Content validity	Measurement model
ComQol-ID (Cummins 1997)	C	A-C [a]	C	ND	ND	A	B	ND
QOLQ (Schalock and Keith 1993)	A-C [a]	A	ND	B	B	A	B	ND
CQ (Stancliffe and Parmenter 1999)	A	ND	B	ND	A	A	B	A
HSIS (Ruddick and Oliver 2005)	A-C [a]	B-C [a]	ND	ND	ND	ND	A	ND
LSS (Heal and Chadsey-Rusch 1985)	A-C [a]	A-B [a]	ND	ND	ND	A	B	ND

Table 6

Summary of the Measurement Properties (Continued)

MLSS (Harner and Heal 1993)	A-C [a]	A-C [a]	ND	ND	C	A	B	ND
PWI-ID (Cummins and Lau 2005)	B	B	ND	B	ND	B	B	ND
MAMP (Bonham et al. 2004)	B	ND	ND	ND	ND	ND	A	ND
POS (van Loon et al. 2009)	B-C [a]	ND	A-B [a]	ND	B-C [a]	A	A	ND

Note. The detail meaning of A, B, and C can refer to Table 1; [a] Range are presented in which reliability information are located within two or three evaluation criteria; ND = No Data
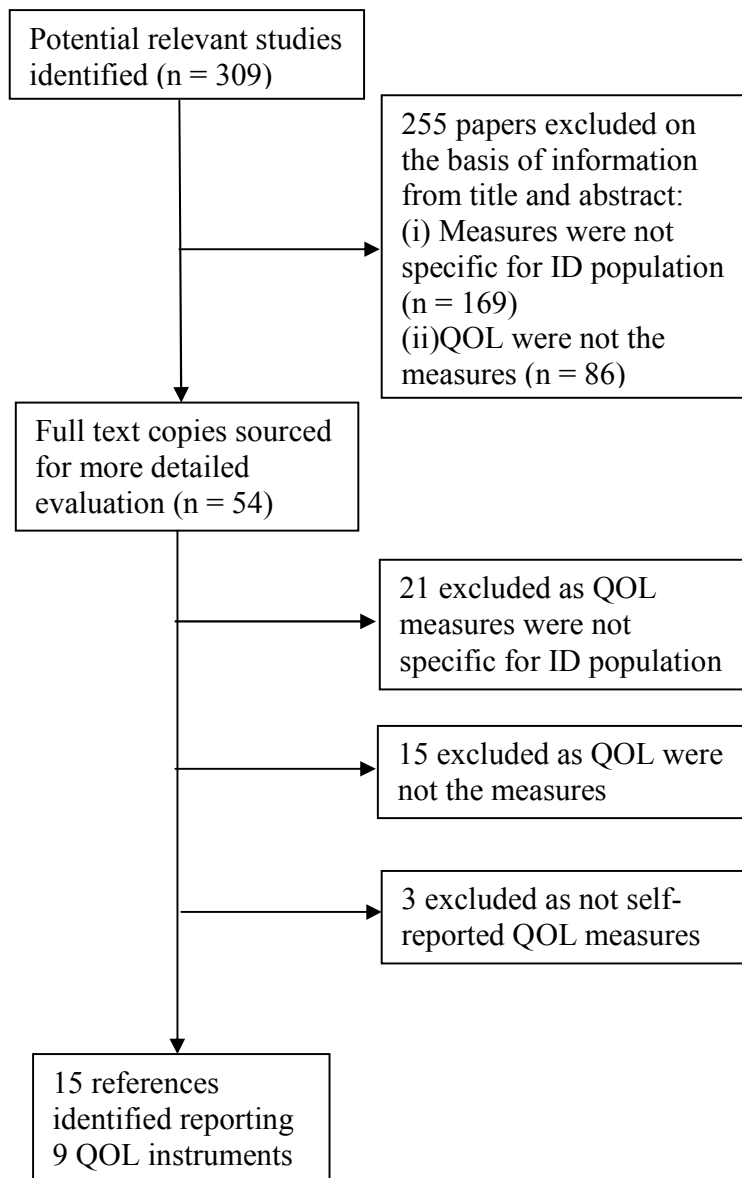
Fig 1   Flow Diagram of Studies Selection of QOL Measures in ID Population