

Strengths and Difficulties Questionnaire (SDQ) in Child and Adolescent Assessment: A Critical Review of Psychometric Evidence, Cross-Cultural Applicability, and Implications for Kenya

Florence A. Ochanda, Dr. Jasper Isoe

Institute of Youth Studies, Tangaza University, Nairobi, Kenya

DOI: <https://doi.org/10.47772/IJRISS.2026.100800409>

Received: 21 August 2026; Accepted: 25 August 2026; Published: 07 September 2026

ABSTRACT

The Strengths and Difficulties Questionnaire (SDQ) is a widely used behavioral screening instrument for assessing emotional and behavioral difficulties among children and adolescents. This is due to its accessibility, brevity and applicability across educational, clinical and research settings. Developed by Goodman in 1997, the SDQ is a 25-item questionnaire comprising five domains: Emotional symptoms, Conduct problems, Hyperactivity/Inattention, Peer Problems and Pro-Social Behavior. The analytical narrative review critically examines the historical development theoretical foundations, psychometric properties, cultural adaptation, clinical utility, and evidence regarding SDQ across global, African and Kenyan contexts. Relevant literature published from 2022 onwards was reviewed, and only seminal studies retained where necessary to provide theoretical and historical context. Literature addressing SDQ reliability, validity, factor structure, measurement invariance, standardization, cultural adaptation and clinical utility was synthesized, with particular attention to evidence from Kenya and other African settings. Overall, the evidence supports the usefulness of the SDQ as a screening instrument although psychometric performance varies across subscales, populations, informants and cultural contexts. Evidence from African settings indicate variability in cultural measurement, performance and factor structure highlighting the importance of context specific validation. In Kenya SDQ has been used increasingly in child and adolescent research, but evidence remains limited particularly regarding measurement invariance, comprehensive validation and locally driven norms. The review supports continued use of SDQ as a screening measure with cautious context sensitive interpretation while emphasizing need for comprehensive Kenyan psychometric validation, cultural adaptation and development of locally appropriate normative data are thus recommended to strengthen validity, applicability and reliability.

Key Words: Strengths and Difficulties Questionnaire, SDQ, Psychometric properties, behavioral assessment, Children, Adolescents, Cultural adaptation

BACKGROUND

Child and Adolescent Psychological Assessment

Childhood and adolescence represent critical developmental periods characterized by rapid emotional, cognitive and social changes. According to the World Health Organization (WHO, 2024) childhood extends from birth to the onset of adolescence, while, adolescence encompasses the age between 10-19 years. The developmental period is rife with diverse biological, environmental and psychological influences that shape both mental and psychosocial adjustment. Evidence worldwide increasingly indicates that approximately one in seven adolescents experience emotional, behavioral and mental health difficulties, stressing the importance of early identification and intervention (WHO, 2024). Equally, valid and reliable assessment approaches that facilitate accurate understanding of developmental strengths and difficulties in child and adolescent psychology cannot be underestimated.

Wright et al. (2022) define psychological assessment as an evidence based clinical process that integrates data from multiple sources, including standardized tests, behavioral observations, interviews and collateral information to answer referral questions, inform diagnosis, inform treatment planning and evaluate outcomes.

In child and adolescent psychology, assessment is important and it serves multiple functions including screening, diagnosis, treatment planning, progress monitoring and evaluating treatment outcomes. Effective assessment permits educators, clinicians and researchers implement evidence-based interventions, identification of developmental concerns that promote well-being functioning and adaptive functioning. Contemporary evidence suggests that assessments emphasize multidimensional and strength-based perspectives that recognize children's competencies alongside psychological difficulties (Younstrom et al., 2022). Therefore, in child and adolescent mental health, assessment constitutes a cornerstone of evidence practice.

Standardized psychological assessment tools globally have become indispensable in educational, clinical, and research settings. Instruments include Strengths and Difficulties Questionnaire (SDQ) developed by Robert Goodman in 1997, Child Behavior Checklist (CBCL) developed by Thomas M. Achenbach in 1983 with major revisions in 2001 and 2014, Corners Rating Scale and Behaviour Assessment System for Children (BASC-3) was developed by Cecil R. Reynolds and Randy W. Kamphaus in 2015 are widely employed to screen and evaluate behavioral and emotional problems among child and adolescents. SDQ was developed by Goodman in 1997 has gained international recognition because of its accessibility, brevity and ability to assess both prosocial strengths and psychopathology.

Studies support psychometric utility of the SDQ has been emphasized across Europe, Asia and Australia emphasizing the importance of examining its factorial validity and measurement invariance across diverse populations (Theunissen et al., 2024), A similar study carried out in Australia in screening adolescents with behavioral and emotional disorders reported satisfactory validity of the SDQ, confirming that it is useful when used as population-based screening instrument (Sawyer et al., 2024). Therefore, global reliance on standardized measure in child and adolescent counselling was underscored by the findings.

Among the African population, the increased awareness of child and adolescent mental health assessment has expanded psychological assessment practices. Persistent concerns emerge when using assessment instruments developed in the Western contexts raising questions of cross-cultural applicability. Contextual variations, linguistic diversity and cultural differences may influence the validity and interpretation of assessment outcomes (Youngstrom et al., 2022). The use of assessment SDQ instruments across multiple countries has been observed, especially in South Africa, while equally stressing the challenges relating to translation, inadequate challenges relating to translation and cultural adaptation (Aarø, 2022). Despite African scholars increasingly advocating for culturally responsive assessment practices, limited empirical validation studies still remains limited. Before adopting standardized- instruments for diverse populations there is need to evaluate measurement equivalence and context appropriateness, as have been observed in the recent development in African child and adolescent mental health research (Klasen et al., 2022). Therefore, the priority for this contentment is to strengthen psychometric foundations of assessment tools.

In Kenya, the increasing demand for mental health services, educational assessments, and developmental screening for children and adolescent has amplified the demand for psychological assessment and this remains an emerging field. The growing prevalence of behavioral and emotional difficulties among children and adolescents has heightened the need for standardized evidence based-assessment tools (Basheer, et al., 2025). Challenges within psychological assessment practices in Kenya are extensive. These include limited availability of culturally validated instruments, inadequate local norms, reliance on measures standardized in high income countries, limited availability of culturally validated instruments and shortage of trained professionals. Subsequently, psychometric adequacy and cultural sensitivity remain important considerations for practitioners and researchers. Current developments in education psychology and mental health assessment practices in Kenya underscore the need to improve diagnostic accuracy and strengthen evidence-based intervention and planning.

The significance of this paper lies in its role in contributing to evidence-based child and adolescent psychology. With the increase use of standardized instruments, this necessitates critical evaluation of validity, reliability, practical utility before application and cultural appropriateness before application. Assessment tools do not function uniformly across populations; hence assumptions must be interrogated due to

socioeconomic and cultural linguistic contexts which may influence responses and interpretation. Through evaluating the Strengths and Difficulties questionnaire, accurate identification of emotional and behavioral difficulties is possible in order to prevent misclassification or inappropriate interventions (Groth-Marnat & Wright, 2024). There is need for rigorous evaluation which contributes to strengthening scientific foundation and development of culturally sensitive assessment practices in child and adolescent mental health services.

To examine the psychometric evidence, cross cultural applicability and clinical utility of the Strengths and Difficulties Questionnaire (SDQ), this paper adopts a critical narrative review design, with particular attention to evidence from Kenya and other African contexts. A critical narrative approach was selected because diverse empirical studies can be examined, compared and interpreted across different populations. The focus of this review is on reliability, validity, factor structure, standardization, cultural adaptation and critical utility of the SDQ instrument and uses synthesized evidence to identify gaps in its application and psychometric validation.

Statement of the Problem

The basis of evidence-based practice in child and adolescent psychology is psychological assessment. Standardized assessment tools should ideally demonstrate strong psychometric properties, including validity, measurement invariance, reliability, and responsiveness across diverse populations. The tools should also exhibit linguistic and cultural appropriateness to ensure that assessment outcomes accurately reflect children's behavioral and emotional functioning within their cultural and social contexts (Groth-Marnat & Wright, 2024). According to the International testing standards, assessment tools must undergo rigorous validation processes whenever they are introduced to a new cultural setting rather than assuming evidence obtained in one population can be generalized to another context (International Test Commission, (ITC, 2023). Psychometrically robust and culturally appropriate measures improve accuracy of screening, diagnosis, intervention planning and programme evaluation thus reducing measurement error and diagnostic bias.

However, in many low- and middle-income countries including Kenya, the reality differs considerably from this ideal situation. Most standardized assessment instruments including the Strengths and Difficulties questionnaires (SDQ), were developed and normed within the Western contexts before being adopted in the African settings.

The use of assessment instruments inadequately validated in the Kenyan context has important implications for child and adolescent mental health services. Assessment results may reflect cultural differences in behavior expressions rather than genuine psychological difficulties without evidence of cultural evidence. Children may be hence be misclassified, leading to overidentification or under identification of behavioral and emotional problems. This often leads to inaccurate diagnoses and inappropriate intervention planning (Groth-Marnat & Wright, 2024). Such limitations compromise research quality, educational decision making and allocation of mental health resources. Advocates of contemporary psychometric research, indicate the need for continuous evaluation of standardized instruments to ensure fairness, validity, and applicability across diverse population (ITC, 2023).

In the Kenyan context evidence comprehensively evaluating its psychometric properties and contextual properties and contextual applicability remains limited within educational and mental health research; with evidence comprehensively evaluating its context applicability and psychometric quality largely limited. Specifically, insufficient local evidence regarding construct validity, reliability, cultural sensitivity and measurement invariance among Kenyan children and adolescent. The gap constrains evidence-based psychological assessments and interventions while limiting confidence in the interpretation of SDQ scores. Therefore, this paper seeks to critically examine the psychometric quality, evidence of validation and contextual applicability of SDQ for use in the Kenyan context. The review seeks to determine whether SDQ provides an accurate and contextually relevant measure of emotional and behavioral functioning. Synthesizing current evidence on its reliability, validity and cultural relevance among Kenyan children and adolescents will help identify properties for future psychometric research.

METHODOLOGY

Review Design

To critically examine the psychometric properties, the paper employed a narrative review design examining psychometric properties, cross-cultural applicability, and clinical utility of the Strengths and Difficulties Questionnaire (SDQ) with specific attention to evidence from Kenya and other African contexts. A narrative approach was employed because the literature includes studies with diverse population, SDQ variations, study designs and psychometric outcomes that are not comparable through quantitative synthesis. The review focused specifically on literature addressing the reliability, validity, measurement invariance, factor structure, cultural adaptation, standardization and clinical utility of the SDQ among children and adolescents. Emphasis was particularly placed on studies published from 2022 onwards to capture contemporary psychometric evidence, while seminal studies were retained to describe the development and original validation of the instrument.

Literature Search

A literature search was conducted using Google scholar, Web of Science, PubMed, PsycINFO, Scopus. Search terms included combination of words such as “Strengths and Difficulties Questionnaire,” “Psychometric properties,” “validity,” “Reliability,” “measurement invariance,” “factor structure,” “cultural adaptation,” “children,” “Kenya,” “adolescents,” and “Africa,” Boolean operators AND and OR were used to search for terms

Eligibility Criteria

Studies eligibility criteria included; reported evidence related to validity, reliability, measurement invariance, factor structure, standardization clinical utility, or cultural adaptation, examined the SDQ among children and adolescents, included populations from Kenya, Africa or relevant international contexts, were published in English and were published from 2022 onwards, except seminal studies retained for theoretical and historical contexts.

Studies that focused on adults, did not examine SDQ, or did not provide relevant psychometric assessment were excluded.

Data Synthesis

Information that was relevant was extracted and synthesized thematically according to major psychometric domains of validity, reliability, norm development, standardization, clinical utility and cultural adaptation. Due attention was given to differences in findings across informant versions, populations, and cultural settings. Kenyan and African evidence was considered separately to identify the extent to which international psychometric findings can reasonably be applicable in the Kenyan context.

CONCEPTUAL AND THEORETICAL PERSPECTIVES

Conceptual Analysis

Psychological Assessment

Psychological assessment is the systematic process of collecting, integrating and interpreting information about an individual's emotional, cognitive, social, behavioral and developmental functioning. This is done through the use of multiple sources of evidence and standardized Instruments. Assessment extends beyond diagnosis to include screening, case conceptualization, intervention planning, progress monitoring and outcome evaluation in child and adolescent Psychology (Groth-Marnat & Wright, 2024). Clinicians in contemporary assessment practices emphasize evidence-based decision making through integration of psychometric evidence with developmental history, caregiver reports, teacher reports, behavioural

observations and clinical interviews with the aim of understanding children's strength and psychosocial difficulties comprehensively. No single assessment instrument can be used in isolation when making clinical or educational decisions (Mash & Hunsley 2023).

The SDQ, facilitates timely recognition of psychosocial concerns while simultaneously identifying the children's strength, thus supporting a strengths-based approach intervention. Recent scholarship has shifted psychological assessment towards a holistic evolution that considers developmental competencies along emotional and behavioral challenges form a deficit-oriented model. Evidence-based assessments encourage clinicians to use instruments with contextual and psychometric relevance rather than relying on convenience or tradition (Havisto et al., 2023). This shift is important particularly in multi-cultural societies where expression of psychological distress and behavioral norms vary across cultural contexts. Subsequently, the selection of assessment instruments should rely on empirical evidence regarding their reliability, validity, clinical utility and cultural appropriateness.

Psychometric Properties

This refers to the scientific characteristics that determine the credibility and quality of psychological assessment instruments. These properties provide empirical evidence regarding an instruments ability to measure the intended construct consistently and accurately (Groth-Marnat & Wright, 2024). Contemporary psychometric evaluation extends to examining construct validity, measurement invariance, responsiveness, and fairness across diverse populations and not only limited to reporting reliability coefficients (Mokkink et al., 2024). SDQ's psychometric evaluation determines whether the questions provide meaningful interpretation information about the children's emotional and behavioural functioning. As instruments are applied through different populations, International Testing Standards emphasized that psychometric evidence should be generated continuously rather than assuming that evidence obtained in one setting can be generalized automatically to another setting (ITC, 2023).

Reliability

Reliability refers to the degree to which an assessment instrument consistently produces dependable scores across repeated administrations or measurement conditions, thus reducing measurement error and supporting accurate clinical interpretation (Wirght et al., 2022). In the original validation of SDQ, Goodman (2001) examined its reliability using a nationally representative sample of 10,438 British children and adolescents aged between 5-15 years. Findings demonstrated acceptable psychometric performance across parent -, teacher-, and self-report versions of the instrument. Using Cronbach's alpha, the internal consistency averaged $\alpha = .73$, indicating acceptable homogeneity among the SDQ test items. Over a period of four to six months, Test-retest reliability produced a mean coefficient of $r = .34$ reflecting moderate agreement across the respondents (Goodman, 2001). The Total Difficulties Score consistently demonstrates stronger reliability than some individual subscales mainly conduct problems and peer problems which tend to produce lower coefficients et al., 2023). SDQ is overall a reliable behavioral screening instrument for identifying both emotional and behavioral difficulties among children and adolescents. Reliability should however be re-evaluated whenever an instrument is culturally adapted, translated or applied to new populations to ensure consistent measures across different contexts (ITC, 2023).

Validity

The extent to which an instrument accurately measures the construct it is intended to measure is referred to as validity. Goodman (2001) demonstrated satisfactory evidence of construct, criterion and discriminative validity across the parent-, teacher-, and self-report versions during the original validation. Construct validity was significantly supported by associations between the SDQ subscales and the intended dimensions of emotional and behavioral functioning. Criterion validity was established through comparisons with the Development and Well-Being Assessment (DAWBA) and clinical diagnoses, showing that higher SDQ scores were significantly associated with an increased possibility of psychiatric disorders (Goodman, 2001). Evidence supporting SDQ equally demonstrated good discriminant validity successfully distinguishing clinical form community samples, with analysis from Receiver Operating Characteristics (ROC) yielding

areas under the curve (AUCs) ranging from 0.84 to 0.94, indicating good to excellent diagnostic accuracy (Goodman et al., 2000). Additionally, content validity was ensured during instrument development through literature review and expert consultations, ensuring comprehensive coverage of key domains of child and adolescent mental health (Goodman, 1997). These findings, collectively support the SDQ as a valid behavioral screening instrument.

Standardization

Standardization is the uniform administration, scoring, interpretation and normative comparison of assessment instruments. Standardization procedures ensure that differences obtained in score reflect actual differences in psychological functioning rather than inconsistencies in testing conditions and aim to reduce bias (Cohen et al., 2022).

SDQ was standardized in the United Kingdom using large community and clinical samples to ensure consistent administration, scoring and interpretations across different populations (Goodman, 1997, Goodman, 2001). The instrument comprises 25 items scored on a three-point Likert scale (0 = Not true, 1 = Somewhat True, and 2 = Certainly True). Scores from Emotional Symptoms, Conduct Problems, Hyperactivity/Inactivity, and Peer Subscales are summed to produce a Total Difficulties Score ranging from 0-40, while Prosocial Behavior scale is interpreted separately (Goodman, 2001). Normative cut-off scores classify results into Normal (approximately 80% of the population), Borderline (10%), and Abnormal (10%) with categories based on UK population norms. Clinicians and researchers can identify children and adolescents who may require psychological assessment using these thresholds. Subsequently SDQ has been translated into 90 languages with standardized administration procedures available in parent-, teacher-, and self-report versions. The ITC (2023) emphasizes on the importance of appropriate adaptation, scoring and interpretation when tests are used across cultural and linguistic contexts. Locally derived norms and cultural adaptations should be established before applying the original UK cut-off scores in a different cultural context. Studies carried out in non-western contexts have shown that SDQ generally retains acceptable screening utility.

Cultural Adaptation

The need for culturally appropriate psychological assessment instruments that yield meaningful and valid interpretations results across diverse population has intensified across the globe. Cultural adaptation involves systematically modifying an assessment instrument to ensure conceptual, semantic, linguistic, and functional equivalence, while at the same time preserving the underlying construct being measured (ITC, 2023). Theunissen et al. (2024) reported that educational level influenced the comprehension of the questionnaire items, highlighting the need for simplified wording and cultural adaptation where necessary. Equally, studies conducted in non-western settings have shown that modifications to item wording, scoring procedures or scoring may improve measurement accuracy within local populations despite SDQ generally retaining acceptable screening utility. Research evidence suggests that SDQ continues to demonstrate satisfactory overall psychometric performance as a behavioral screening instrument. However recent validation studies have shifted attention towards examining structural validity, cultural adaptation, measurement invariance, and contextual applicability. Therefore, psychometric adequacy should be established through rigorous validation, particularly in linguistically and culturally diverse settings such as Kenya, where nationally representative norms and comprehensive validation studies remain limited.

Clinical Utility

Clinical utility refers to the extent to which an instrument contributes meaningfully to assessment, diagnosis, intervention planning, progress monitoring, and clinical decision making (Hunsley & Mash, 2023).

Developed by Goodman (2002), SDQ has gained widespread acceptance due to its favorable and brief psychometric properties, with recent studies indicating that its factor structure, measurement invariance and score interpretation- differing across linguistic and cultural groups (Theunissen et al., 2024; Sheel et al., 2025). The findings highlight the need for context specific validation before widespread educational or clinical application. De Los Reys et al., (2024) contends that multiple sources of information should be integrated to

inform behavioral manifestations that differ across home and school environments. Youngstrom et al. (2023) contends that standardized questionnaires should not replace comprehensive clinical assessment but rather complement standardized questionnaires. Accordingly, the usefulness of the SDQ within Kenya does not only depend upon its cultural relevance but also on psychometric properties and its capacity to inform educational and mental health interventions.

Table 1. Selected Psychometric Evidence on the Strength and Difficulties Questionnaire in Kenya and Africa

Country	Study	Sample (N)	Key Psychometric Finding	Implication
Kenya	Mbithi et al. (2023)	797	25 items SDQ $\alpha = .75$	Supports internal consistency, but not comprehensive validation
Kenya	Baseke et al. (2026)	2,843	10-item SDQ $\alpha = .582$ for	Suggests caution when using abbreviated versions
South Africa	Aarø et al. (2022)	3,542 adolescents	Original five factor model showed poor fit; alternative-structure provided better fit.	Supports context specific examination of factor structure
Democratic Republic of Congo DRC	Kashala et al. (2005)	1,306 children	Examined SDQ factor structure and cutoffs	Illustrates the importance of local validation
Africa	Hoosen et al. (2018)	54 studies multiple African populations	Widespread use, but substantial variation in methodological and psychometric evidence	Highlights broader African validation gap

Theoretical Foundations of Strengths and Difficulties Questionnaires

Developmental Psychopathology Framework

The strengths and difficulties Questionnaires (SDQ) were developed by Robert Goodman in 1997 as a brief behavioral screening instrument with the aim of assessing emotional and behavioral functioning among children and adolescents (Goodman, 1997). Conceptually informed by the developmental psychopathology framework which theorizes that children's behavioral and emotional challenges arise from the interaction between psychological, biological and environmental influences across development. SDQ is consistent with this framework assesses biobehavioral difficulties and personal strengths through five domains; emotional symptoms, conduct problems, Hyperactivity/Inattention, peer relationship problems, and prosocial behavior. Assessing both strengths and difficulties therefore provides a balanced understanding of children's, psychosocial functioning and mental health (Goodman, 1997; Goodman, 2001). The psychometric perspective, SDQ was initially developed and validated using Classical Test Theory (CTT). CTT assumes that an individual's observed score consists of a true score and a random measurement error (Crocker & Algina, 1986). CTT principles were used to evaluate item performance, reliability, internal consistency, and construct validity. These psychometric properties were examined by Goodman (1997), to ensure that the questionnaire

reliably measured the intended behavioral and emotional constructs. SDQ has continually been assessed using the Item Response Theory (IRT) and Rasch models to assess cross-cultural validity and measurement invariance (Ortuno-Sierra et al., 2015).

Evidence Based Assessment

Evidence Based Assessment (EBA) integrates psychometric evidence, contextual information and clinical expertise to guide decision-making. EBA advocates for the use of measurement supported by robust evidence of reliability, validity, cultural utility and appropriateness rather than selecting assessment instruments based on familiarity or convenience (Hunsley & Mash, 2023). This approach encourages the integration of standardized questionnaires with developmental history, behavioral observations, interviews and report from multiple informants and recognizes that no single instrument is sufficient for comprehensive psychological assessment. While it offers a framework through which SDQ can be evaluated, its emphasis on contextual interpretation, multidisciplinary decision-making and scientific evidence supports responsible application of standardized screening tools within naturally diverse settings.

CONCEPTUAL FRAMEWORK

The assessment theory guided this review which forms the scientific foundation for evaluating SDQ. Classical Test Theory, Item Response Theory, and Evidence-Based Assessment provide the theoretical basis for examining psychometric properties of instrument, specifically reliability, validity and standardization. Evidence resulting from these properties inform the validation of SDQ by determining its accuracy in measuring children's emotional and behavioral functioning within diverse populations. Validation supports its clinical application and cultural appropriateness in screening, intervention, planning referral and monitoring outcomes. The framework thus conceptualizes assessment theory, psychometric properties, validation, cultural appropriateness and clinical application as interrelated components that determine collectively the suitability of SDQ for use in child and adolescent psychological assessment within the Kenyan context.

Historical Development and Description of the Tool

Origins and Evolution

The Strengths and Difficulties Questionnaire (SDQ) was developed in 1997 by Robert Goodman as a brief behavioral instrument for identifying emotional and behavioral problems among children and adolescents. The SDQ was developed with the aim of overcoming limitations connected to earlier behavioral screening measures especially the cost, length and predominant importance on psychopathology. Goodman sought to produce an instrument that was psychometrically robust, easy to administer, brief and applicable across clinical, educational and research settings recognizing children's strength and their behavioral difficulties (Goodman, 1997). Compared to many earlier instruments SDQ incorporated positive behavioral attributes, thereby providing more balanced assessment of child and adolescent mental health.

The SDQ instrument has undergone substantial development since its introduction to improve its applicability across different age groups and diverse cultural settings (Stone et al., 2015). Goodman subsequently developed corresponding versions (Goodman, 1999) have equally been developed for parents, teachers, and self-report by young people, together with extended versions incorporating an impact supplement that assesses distress, chronicity, impairment and burden on others.

The questionnaire widely used in epidemiological schools, surveys, specialist mental health services and research worldwide has also been translated into more than 90 languages (Goodman, 2001). Its widespread adoption, particularly in low- and middle- income countries is further enhanced due to its availability at no cost for commercial use.

Purpose

SDQ was developed with the aim of rapid behavioral screening assessment for identifying children and adolescents who may be experiencing emotional or behavioral difficulties requiring further evaluation and

intervention (Goodman, 1997). SDQ assists educators, clinicians, and researchers in monitoring behavioral change over time and evaluating intervention outcomes. The questionnaire is equally appropriate for routine clinical assessment, population-based surveys, educational screening and research involving child and adolescent mental health.

Structure and Domains Assessed

SDQ contains- 25 items organized around five sub scales containing five items each. The domains assessed include; Emotional symptoms, conduct problems, Hyperactivity/Inattention, peer relationship problems, and pro-social behavior. The first four domains are summed up to generate Total Difficulties Score, while the first four domains collectively measure psychological difficulties (Goodman, 1997). The Prosocial Behavior subscale is scored separately because rather than reflecting psychopathology it reflects positive social functioning. Responses are recorded using a three-point Likert scale comprising 0 = Not true, 1 = Somewhat True, and 2 = Certainly True. Simultaneous assessment of behavioral strength and difficulties is enabled within a brief administration time of approximately five to ten minutes.

Target Population

The SDQ is intended for children and adolescents between 2-17 years, with different version tailored to the developmental stages. A self-report version is designed for adolescents between 11-17 who are considered capable of evaluating their own emotional and behavioral functioning. Parent teacher questionnaires are available for children aged 4-17 years, while a modified versions available for children between 2-4 years. Behavioral assessment is made possible through the availability of multiple informant versions used across home, school, and community environments, thus providing comprehensive understanding of children's psychosocial functioning (Goodman, 2001).

Administration Procedures

The SDQ is a straightforward instrument designed for administration by teachers, parents, healthcare professionals, and adolescents completing the self-report version (Goodman, 2001). The administration may be completed using paper based or electronic formats and requires five and ten minutes. Standard instructions require respondents to evaluate the child's behavior over the preceding six months although during follow-up assessments the shorter reference version may be used. The instrument is suitable for use community surveys, educational settings, routine clinical practice and requires minimal training.

Scoring Procedures

The SDQ items are scored on a three-point Likert scale ranging from 0 to 2, with several positively worded items reversed and scored before calculating the subscale totals. Scores from the Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, and Peer Relationship Problems are summed to produce a Total Difficulties Score ranging from 0 to 40, whereas the Prosocial Behavior Score is interpreted independently and indicates Positive Social competence. Higher Total Difficulties Scores ranging from 20 – 40 indicates great emotional and behavioral difficulties that may warrant further assessment, while higher prosocial Behavior Score of 6-10 reflects strong prosocial functioning (Goodman, 1997). The additional extended version identifies difficulties that interfere with functioning at home, school and peer relationships, thus enhancing clinical interpretations of screening results.

Psychometric Properties and Validation Studies

Reliability

Globally, studies consistently report acceptable internal consistency for the SDQ Total Difficulties Scale with Cronbach's alpha coefficients generally ranging from 0.73 to 0.84 indicating satisfactory internal consistency for behavioral screening though variability is frequently observed across individual sub scales, particularly conduct problems and peer problems (Liegmann et al., 2025). The emotional symptoms and hyperactivity scales generally demonstrate acceptable coefficients above 0.70, whereas peer problems and conduct problems

subscales frequently report lower alpha values between 0.50 and 0.69, reflecting greater variability across populations (Karlsson et al, 2022). While the results indicate overall good performance, some subscales require cautious interpretation particularly when applied in culturally diverse settings. Recent investigations further demonstrate internal consistency may vary due to educational level, age, language and cultural context indicating that reliability should be interpreted within specific populations and not assumed universally (Theunissen et al., 2024).

Test reliability evaluates temporal assessment score stability. SDQ has demonstrated moderate to high stability over intervals ranging from two weeks to six months across validation studies, with correlation coefficients commonly exceeding 0.70 for the Total Difficulties Scale (Bergstrom & Baviskar, 2020). Thus, findings point to SDQ producing relatively stable behavioral assessments over time while remaining sensitive to genuine developmental and environmental changes

Inter-rater reliability tests indicate low to moderate agreements, with correlation coefficients typically ranging between 0.30 and 0.60, reflecting differences in children's behavior across home and school environment (De Los Reyes et al., 2023). These discrepancies provide valuable contextual information regarding children's functioning in different settings rather than representing measurement error. Therefore, evidence-based assessment recommends improving diagnostic accuracy and intervention planning through integrating reports from multiple informants.

Evidence, indicates that SDQ demonstrates acceptable reliability as a behavior screening instrument. Researchers underscore local psychometric evaluation because reliability estimates vary due to age, educational level, language and cultural context.

Validity

Contemporary psychometric theory conceptualizes validity as a unified construct supported by multiple sources of evidence, including content validity, criterion validity, construct validity and cross-cultural validity which assesses the clarity, relevance, and comprehensive individual items (Terwee et al., 2023). SDQ was developed comprehensively across consultations with psychologists, psychiatrists, behavioral scientists and educators to ensure the instrument captures its 25 items comprehensively capturing emotional symptoms, conduct problems, hyperactivity, peer relationship difficulties and prosocial behavior (Goodman, 2001).

Recent global studies support the original SDQ five factor structure with CFI values exceeding 0.9 and RMSEA values below 0.08, indicating acceptable structural validity although some improvements have been noted improved model fit after modifying selected items to accommodate cultural differences (Karlsson et al., 2022). International studies have reported criterion validity to demonstrate correlation moderate to strong correlations between SDQ scores and established measures such as Development and Well-Being Assessment (DAWBA) and Child Behavior Checklist (CBCL) satisfactory specificity and sensitivity for identifying children at risk of emotional and behavioral disorders SDQ has then be supported as an effective screening instrument while emphasizing recommendations that it should not replace comprehensive clinical assessment rather complement it.

Cross-cultural validity multinational studies for SDQ indicate that the instrument generally demonstrates acceptable cross-cultural validity, however several have shown differential functioning across countries, suggesting behavioral response style and expectations may influence score interpretation (Ortuno-Sierra., 2022; Theunissen et al., 2024). These study findings therefore highlight the importance of conducting validation studies before widespread implementation of SDQ in diverse culturally sensitive contexts such as Kenya. (Goodman, R, 2001), Collectively, evidence suggests that across multiple psychometric domains, SDQ demonstrates satisfactory validity. Scholars indicate that validity is population specific rather than universal necessitating continuous evaluation within different cultural and clinical contexts to ensure accurate interpretation and appropriate application in child and adolescent psychological assessment.

Standardization and Norm Development

Fundamental components of psychometric evaluation include standardization and norm development because they provide the basis for consistent administration, scoring and interpretation of assessment outcomes. Standardization ensures respondents are assessed under uniform conditions, whereas referenced data provides values against which the individual scores can be interpreted (AERA, APA, & NCME, 2014).

Supporting the need for local norm development is necessary as recent studies demonstrate that behavioral norms differ across populations. While using the national representative sample of Dutch sample of children and adolescent between 3-14 years, Theunissen et al., (2022) established that an updated SDQ normative data reported differences in Cutt off scores between Dutch and original United Kingdom population. The author concluded that in order to reduce the likelihood of false positive and false negative identification of emotional and behavioral difficulties, country specific norms will improve classification accuracy. Therefore, standardization of psychological instruments involves interpreting locally derived reference values rather relying exclusively on norms developed in a different cultural context. These results have an important implication for the Kenyan context because, no national representative SDQ normative database currently exists. Subsequently, researchers and clinicians frequently interpret SDQ scores using foreign data despite substantial evidence in culture, language, child rearing practices and educational systems. Such practices increase the possibility of misclassification while reducing the precision of behavioral classification. Establishing local normative data in Kenya will strengthen interpretability, fairness and clinical utility of the SDQ.

Global Validation Studies

Although evidence suggests that SDQ's performance varies across populations and settings, global validation studies consistently support it as a psychometrically sound behavioral screening instrument. Validation research emphasizes the need for continuous psychometric evaluation rather than assuming that the instrument performs identically across cultures.

Psychometric performance seemed to vary across settings, highlighting the importance of context-specific validation. In a Dutch sample comprising 4,053 clinically referred and 962 community children, reported good reliability for Total Difficulties Score (Cronbach's $\alpha = .80$) although conduct problems (Cronbach's $\alpha = .59$) and peer problems (Cronbach's $\alpha = .55$) with subscales showing weaker internal consistency (Vugteveen et al., 2022). Equally, Theunissen et al. (2024) used SDQ to evaluate the psychometric performance among adolescents with different educational levels in the Netherlands. Satisfactory internal consistency (Cronbach's $\alpha > .70$), whereas peer problems (Cronbach's $\alpha = 0.48$) subscale produced comparatively lower reliability, reflecting findings reported in previous international studies. Confirmatory factor analysis score supported the original five factor model for whole sample (CFI= 0.832, TLI= 0.810, RMSEA= 0.052, SRMR=0.060) among adolescent with lower educational attainment scores revealed CFI=0.779, TLI=0.750, RMSEA=0.064, SRMR=0.085) while for higher education group revealed CFI= 0.808, TLI=0.783, RMSEA=0.054, SRMR=0.064.

Measurement invariance analyses exhibited partial scalar invariance across educational groups CFI= 0.798, TLI= 0.771, RMSEA=0.057, SRMR= 0.071. The study results found partial measurement invariance, with evidence suggesting that the five-factor structure performed well among adolescents with lower education attainment. These findings indicate that although RMSEA and SRMR suggested acceptable fit, CFI and TLI remained below the recommended threshold of 0.9, indicating only moderate support for the original five factor model. These results indicate that caution should be exercised when comparing means across populations. Among the poorly performing items were identified as they identified the instrument's weakness. Items with standardized regression weights < 3.0 included item 11 ("i have one good friend or more") and item 22 ("I take things that are not mine"). Among the low educated adolescents two additional items showed weak loadings: item 3 ("I get a lot of headaches, stomach aches or sickness") and item 7 ("I usually do as much as I am told"). These findings suggest that certain SDQ items may require cultural adaptation or revision to improve validity (Theunissen, 2024).

The total mean differences demonstrated significant behavioral difference between the educational groups with total difficulties; low education: 10.49 ± 5.66 , higher education score: 8.49 ± 4.76 ($p < 0.1$). These results point to low educated adolescents reported significant higher Conduct Problems, Hyperactivity and Peer problems whereas Prosocial Behavior scores were significantly lower (Theunissen, 2024). Khawaja (2020) demonstrated that although the five-factor structure was retained among linguistically and culturally diverse Australian students, several items functioned differently across language groups. In Finland, Kankaanpää et al., (2023) reported limitations in the SDQ psychometric performance. Though a revised model showed excellent fit among early adolescents, none of the models tested demonstrated acceptable fit among mid adolescents with subscales showing low reliability.

Jointly, the research findings indicate that while SDQ demonstrates construct validity and reliability internationally, local psychometric validation and cultural adaptation remain essential before implementation in new contexts such as Kenya.

African Validation Studies

The contextual applicability of SDQ has been validated in the African setting and have contributed to evidence regarding the instrument demonstrating that psychometric performance is influenced by linguistic, cultural and contextual factors. African research increasingly suggests that the dimensionality of emotional and behavioral problems may differ across populations unlike many European studies that support the original five factor structure, highlighting the importance of local structural validation rather than assuming universal applicability. A comprehensive contemporary validation study was carried out among 3,542 adolescents in Western Cape, South Africa. Hansel et al. (2018) indicated that the original five factor structure model demonstrated poor fit. Four poorly performing items were removed and a three-factor model comprising Internalizing problems, Externalizing problems and Prosocial behavior achieved acceptable model fit (CFI=.921, TLI=.920, RMSEA=.024) demonstrating strong measurement invariance across sex and three language groups (English, Africans, isiXhosa). Thus, broader behavioral dimensions better capture adolescent mental health within multi-cultural African context than the original SDQ structure. Relatively few studies have undertaken comprehensive psychometric validation despite SDQ being widely applied across African countries (Hanse et al., 2018). The scoping review highlighted methodological inconsistencies across studies, including sampling procedures, respondent versions thus limiting comparisons. The usefulness of the SDQ as a behavioral screening instrument is not the principal challenge but, the assumption that its original psychometric structure is universally applicable. This reinforces the importance of international recommendations that standardized psychological measures should undergo rigorous structural validation, local norm development and rigorous cultural adaptation before implementation in diverse African settings including Kenya.

Evidence from Kenya

SDQ use in Kenya has increased considerably especially while examining child and adolescent mental health, psychosocial well-being, HIV related outcomes and psychological effects of COVID 19. Evidence supporting its psychometric adequacy within the Kenyan context remains limited despite its widespread use. Primarily SDQ has been employed as screening instrument rather than comprehensively evaluating its measurement properties thereby creating a significant gap between its empirical evaluation and application.

No published Kenyan study evidence has comprehensively examined the SDQ's structure using measurement invariance, confirmatory factor analysis, Kenya's cultural and linguistic groups, clinical diagnoses, criterion validity or development of normative data. The absence of such evidence thus limits confidence in the interpretation of SDQ scores and raises concerns regarding direct application of cutoff scores derived from western population.

Future research should give priority to rigorous validation involving linguistically and culturally diverse Kenyan population establish local norms, criterion validity, structural validity and measurement invariance thus strengthening clinical and scientific applicability of the SDQ within the Kenyan setting.

Critical Evaluations and Application of Strength and Difficulties Questionnaire (SDQ) in Kenya

Strengths of the SDQ

The SDQ instrument is among the mostly used tool for screening emotional and behavioral difficulties adolescents and children because of its accessibility, brevity and sound psychometric foundation. SDQ assesses both behavioral difficulties and prosocial strengths unlike many behavioral measures that focus primarily on deficits thereby supporting strength-based assessment and intervention planning (Goodman, (Goodman, R, 1999)2001). The multi-informant assessment available for parent-, teacher-, and self-report versions facilitates clinicians, and educators to evaluate children's behavior across different contexts, Validation studies continue to support SDQ effectiveness in screening behavioral and emotional difficulties in community, educational and healthcare settings (Theunissen et al., 2024). Furthermore, its low-resource settings and brevity make it especially suitable for low-resource settings where specialist psychological services remain limited (Hunsley & Mash, 2023).

Limitations of the SDQ

SDQ has important limitations despite its advantages. Psychometric performance varies across languages, populations, culture indicating that findings cannot be automatically generalized to another. For the peer problems and conduct problems, studies consistently report low reliability coefficients than for the Total Difficulties Scale, suggesting variability in how behavioral constructs are interpreted across settings (Theunissen et al., 2024). Besides, SDQ is more a screening rather than a diagnostic instrument. Therefore, assessment findings should be interpreted alongside clinical interviews, behavioral observations, developmental history, and reports from multiple informants to help improve diagnostic accuracy and intervention planning (Aadland et al., 2022).

Cultural Considerations

The application of SDQ in Kenya entails careful attention of the cultural factors that may influence children's symptom interpretation, behavioral expression and assessment outcomes. SDQ has demonstrated satisfactory psychometric properties in most countries, psychological assessment is largely influenced by cultural context in which the behavior is observed and interpreted. Instruments therefore developed in one cultural setting cannot be assumed to possess equivalent reliability and validity in another without empirical evidence supporting cultural applicability (ITC (2023), 2023).

Kenya being a culturally and linguistically a diverse country comprising more than 40 ethnic communities with very distinct traditions, languages, parenting practices and social norms. These cultural differences influence how behavioral and emotional difficulties are perceived by parents, teachers and the community as a whole. For example, behaviors interpreted as independence, hyperactivity or assertiveness in Kenya may be viewed differently from the Western context where cultural expectations surrounding respect, obedience, respect for authority and collective social responsibility are highly valued. Similarly, expressions of emotional distress may be shaped by family practices, community norms and cultural beliefs potentially influence how respondents interpret and answer the items (Baidoo-Anu et al., 2024). When internationally developed norms are applied without validation, such differences may increase the likelihood of potential cultural bias. Establishing semantic, conceptual and functional equivalence ensures that the questionnaire items do not measure the cultural difference in language and behavioral expectation but rather the intended psychological constructs. Within the Kenyan context limited published evidence remains that examines the cultural validity of SDQ across linguistic and cultural groups. Therefore, future research should prioritize development of Kenyan normative data, translation into major local languages and evaluation of measurement invariance across diverse populations. Thus, in order to strengthen the cultural relevance and acceptability of the instrument collaboration among psychologists, educators, linguists and community stakeholders is necessary. Finally, a culturally responsive SDQ adopted tool would improve accuracy of behavioral screening, support equity, reduce risk of misclassification and support evidence-based assessment practices among children and adolescents in Kenya.

Linguistic Challenges

Culture and language present with significant challenges in the application of SDQ. Translation alone does not ensure conceptual evidence despite the questionnaire having being translated into numerous languages. Behavioral concepts such as peer relationships, emotional distress or hyperactivity may be understood differently across cultures, potentially influencing responses (IACT, 2023). A systematic literature review carried out in South Africa indicated that several studies relied on translated versions without comprehensive locally derived normative data or comprehensive psychometric validation, increasing the possibility of inaccurate interpretation and cultural bias (Hoosen et al., 2018). The findings highlight the importance of rigorous cognitive interviewing, back translation and local validation before routine implementation in the Kenyan context.

Ethical Considerations

The ethical psychological assessment requires that instruments demonstrate acceptable reliability, validity, fairness and cultural appropriateness for the population which they are use (AERA, APA, & NCME, 2014). Using inadequately validated instruments may result in stigmatization, inappropriate referrals, misclassification and ineffective intervention and planning. Ethical practice further requires confidentiality, informed consent and interpretation by appropriately trained professionals. Within the Kenyan context practitioners should exclusively avoid relying on international normative data as behavioral expectations are not similar across board but are shaped by local educational, local cultural and family contexts.

Suitability for Kenyan Children and Adolescents

The SDQ has considerably potential to be used in the African context because it is easy to administer, inexpensive and suitable for schools, healthcare facilities and community mental health programs. These characters align with national efforts targeted towards early identification- of both behavioral and emotional difficulties among a children and adolescents. However empirical evidence supporting SDQ among the Kenyan population still remains limited. Hoosen, et al., 2018) posits that existing African reviews including Kenya consistently identify the absence of locally validated norms and comprehensive psychometric studies. Consequently, the need for further validation before the adoption of SDQ as a routine screening instrument within Kenya educational and clinical settings.

Potential Adaptation

SDQ should prioritize translation into major Kenyan languages using internationally recommended procedures, including back-translation, forward translation, expert review, and cognitive interviewing (ITC, 2023). Subsequent studies should examine measurement invariance across different linguistic, cultural, socioeconomic and geographic groups. Such Adaptations would enhance interpretability, fairness and diagnostic accuracy of the instrument within Kenya.

Contemporary and Emerging Issues

Recent development place emphasis on culturally responsive evidence based and digitally supported assessment practices. Increased attention has been given to- locally derived normative data, measurement invariance and screening platforms that improve accessibility and efficiency (Mokkink et al., 2024). These developments thus present opportunities to integrate SDQ into school mental health programs and healthcare services while ensuring implementation is supported by robust local psychometric evidence. Current psychological assessment increasingly underscores inclusive assessment, technological innovation and cultural responsiveness. Inclusive assessments promote the use of accessible, equitable and appropriate instruments for children and adolescents form diverse linguistic, cultural and socio-economic backgrounds, thus reducing assessment bias while equally improving fairness (ITC (2023), 2023). This also includes advances in digital assessments which includes electronic screening, telepsychology and electronic screening enhancing efficiency and accessibility of child mental health assessment (APA, 2023). Innovations require

rigorous psychometric evaluations to ensure they maintain reliability, validity and cultural appropriateness (Mokkink et al., 2024).

Future research should focus on developing culturally appropriate norms, local normative data, validation of digital versions of SDQ and digital assessment systems that strengthen child and adolescent mental health assessment without compromising scientific rigor.

Practical Implications

Clinical Practice

SDQ, instrument is widely recognized as a reliable behavioral screening tool in child and adolescent mental health. The multidimensional structure enables clinicians to identify conduct problems, emotional symptoms, peer relationship difficulties, hyperactivity and prosocial behavior., thereby facilitating early detection of psychological difficulties and timely interventions (Goodman, 2001). Evidence based assessment advocates integrating standardized measures such as SDQ with behavioral observations, clinical interviews, developmental history and multi-informant reports to help improve diagnostic accuracy and treatment planning (Youngstrom et al., 2023). Evaluation studies have recently further demonstrated that SDQ is effective for monitoring behavioral change over time and evaluation outcomes making it a valuable instrument for child and adolescent clinical services (Theunissen et al., 2024). Behavioral screening tools are designed to identify risk rather than establish diagnosis researchers therefore consistently indicate that SDQ should complement rather than replace comprehensive clinical assessment (Hunsley & Mash, 2023).

Educational Assessment

SDQ serves as an effective screening tool within the education setting for identifying learners experiencing emotional, behavioral and even social difficulties that may affect school adjustment and educational achievement. Teachers can use this instrument to identify students who need additional behavioral interventions, prosocial support or referred for specialized services. Research indicates that both emotional, behavioral difficulties and are associated closely with school absenteeism, poorer educational outcomes and reduced educational attainment, highlighting the importance of early identification within schools (Dos Los Reyes., 2024). SDQ supports early identification of learners requiring prosocial support leading to development of targeted educational interventions. The instrument further aligns with inclusive education for facilitating early recognition of learners with socio-emotional needs before difficulties become severe. Nevertheless, assessment findings should be interpreted alongside classroom academic performance, contextual information and observation. In order to minimize inappropriate labelling while ensuring equitable educational decision making ((AERA, APA, & NCME, 2014).

Counselling Practice

The SDQ provides a structured framework for assessing children's emotional strengths and behavioral difficulties for school counsellors and mental health professionals. The use of assessment assist counsellors in identifying priority areas for establishing counselling goals, intervention, monitoring therapeutic goals and evaluating intervention effectiveness. Inclusion of prosocial behavior scale enables practitioners to adopt a strength-based approach that recognizes both protective factors and areas of concern (Goodman, 2001). Integrating standardized assessment findings with family interviews, teacher consultations and psychosocial assessments to provide holistic and culturally responsive interventions has been emphasized within contemporary counselling practice (Youngstrom et al., 2023). Within Kenyan multicultural context, counsellors should consider family dynamics, community influences, cultural values when interpreting SDQ score to ensure culturally appropriate practice.

Research Application

The SDQ possess sound psychometric properties, brevity and international applicability hence one of the most extensively used instruments in child and adolescent counselling. Researchers employ the instrument in school

mental health surveys, epidemiology studies, Longitudinal investigations and intervention evaluations in examining the prevalence and developmental trajectories of emotional and behavioral problems (Theunissen et al., 2024). Its applicability globally facilitates cross-national comparison and contributes to accumulation of evidence regarding child mental health. Contemporary international research emphasizes meaningful comparisons require evidence of cross-cultural validity and measurement invariance before findings can be generalized across populations (ITC (2023), 2023). Within the Kenyan context, further research is necessary to evaluate structural validity, establish locally validated norms, and determine whether SDQ performs consistently across diverse cultural and linguistic groups.

Policy Implications

SDQ supports evidence-informed planning, early identification, and preventive intervention, therefore it has a significant implication for educational and child mental health policy. Standardized behavioral screening can assist policy makers in estimating the prevalence of behavioral and emotional difficulties, identifying vulnerable populations, allocating mental health resources, and evaluating school-based intervention programs. Current literature increasingly advocates integrating validated psychological screening into comprehensive school mental health frameworks in order to strengthen prevention- and develop early intervention strategies (Margaretha et al., 2023). The implementation of the SDQ within Kenya could contribute to national initiatives promoting inclusive education, child mental health and psychological well-being. Policy implementation should be accompanied by rigorous culturally appropriate normative data, local validation, professional training and ethical assessment guidelines to ensure equitable service delivery and accurate interpretation (AERA et al., 2014; ITC (2023), 2023).

RECOMMENDATIONS

Rather than relying primarily on evidence generated in other populations, future research should undertake a comprehensive psychometric validation of the SDQ among Kenyan children and adolescents. Validation studies should ideally first establish content validity and cultural relevance of the SDQ through cognitive interviewing, expert review and engagement with the relevant stakeholders including children, teachers, parents and mental health professionals. This will help determine whether items are relevant, comprehensible and representative of behavioral and emotional functioning within the Kenyan context. Existing measuring guidance place emphasis on the importance of establishing content validity and internal structure before interpreting measurement properties (

Subsequent research should examine internal structure of the SDQ using exploratory and confirmatory factor analysis to determine whether the original five factor model is supported among Kenyan children and adolescents. Studies should assess test-retest reliability, internal consistency and inter-rater across teacher, parent and adolescent versions. Where appropriate Rasch analysis or Item Response analysis could provide additional evidence regarding item functioning. Different measurement properties require appropriate study statistical analyses and study design (Zhong et al., 2022).

Given Kenya's cultural and linguistic diversity, validation studies should additionally examine measurement invariance across relevant groups including age, language, sex, educational context and geographical location. This would establish whether SDQ scores can be meaningfully compared across these groups rather than assuming that identical scores have equivalent meanings.

Future studies should additionally examine criterion and construct validity by comparing SDQ scores with established psychological and clinical measures and, where appropriate or ethical, clinical assessment. Finally, the large and significant diverse Kenyan samples should be use to develop locally appropriate cut off scores and normative data. This would therefore reduce reliance on norms developed in other populations and improve the fairness and interpretability in Kenya. Interpretation of measurement instruments and evidence-based selection require consideration of reliability, validity and the intended context of use.

Overall, evidence supports SDQ as a useful measure of child and adolescent emotional and behavioral difficulties, although psychometric performance is not uniformly consistent across populations, informants,

subscales and cultural contexts. The key gap is not the absence of validation evidence, but the need for context-specific and methodological rigor in validation, particularly in underrepresented populations.

REFERENCE

1. Aadland, K. N., Lervag, A., Ommundse, Y., & Aadland, E. (2022). Structural Validity of the Norwegian Version of the Strengths and Difficulties Questionnaire in Children Aged 3-6 Years. *Frontiers in Psychology*, 13, Article 1024918.
2. Aarø, L. E., Davids, E. L., Mathews, C. Wubs, A. G., Smith, O. R. F., & de Vries P. J. (2022). Internalizing Problems, Externalizing Problems, and Prosocial Behaviour- Three Dimensions of the Strengths and Difficulties Questionnaire (SDQ): A Study Among South African Adolescents. *Scandinavian Journal of Psychology*, 63(4), 415-425.
3. American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
4. Baseke, R., Kilonzo, R., Ngesa, M., Mwende, P., & Osborn, T. (2026). A Dataset on Adolescent Mental Health in Kenya. *Data Briefs*, 65, 112513.
5. Basheer, S., Tan, S. F., David, R., Majumder, P., & Sayal, K. (2025). The Acceptability and Effectiveness of Standardized Diagnostic Assessment Approaches in Children and Young People's Mental Health Services: An Updated Systematic Review. *Child and Adolescent Mental Health*, 30(4), 364-374.
6. Bergstrom, M., & Baviskar, S. (2020). A Systematic Review on Some Reliability and Validity Issues Regarding the Strengths and Difficulties Questionnaire. *Journal of Evidence-Based Social Work*, 17(6), 1-31.
7. Carneiro, F. A., Costa, P. A., & Leal, I. (2023). Construct-related Validity of Strengths and Difficulties Questionnaire with Three Dimensions: A Multitrait-Multimethod Analysis. *Journal of Health Psychology*, 28(4), 412-425.
8. Cohen, J. R., & Choi, J. W. (2022). Is ACEs Screening for Adolescent Mental Health Accurate and Fair? *Prevention Science*, 23(7), 1216-1229.
9. de Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory*. Guilford Press.
10. De Los Reyes, A., Wnag, M., Lerner, M. D., Makol, B. A., Fitzpatrick, O. M., & Weisz, J. R. (2023). The Operations Triad Model of Youth Mental Health Assessments: Catalyzing a Paradigm Shift in Measurement Validation. *Journal of Clinical Child and Adolescent*, 52(1), 19-54.
11. de Vries, P. J., Davids, E. L., Mathews C., Aaro. L. E., Wubs, A. G., & Smith, O. R. F. (2018). Measuring Adolescent Mental Health Around the Globe: Psychometric Properties of the Self-Report Strengths and Difficulties Questionnaire in South Africa, and Comparison with UK, Australian and Chinese data. *Epidemiology and Psychiatric Sciences*, 27(4), 369-380.
12. DeVellis, R. F., & Thorpe, C. T. (2022). *Scale Development: Theory and Applications*. SAGE Publications.
13. Flake, J. K., Davidson, I. J., Wong O., & Pek, J. (2022). Construct Validity and the Validity of Replication Studies: A Systematic Review. *American Psychologist*, 77(4), 576-588.
14. Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.
15. Goodman, R. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *Journal of Child Psychology and Psychiatry*, 40(5), 791-799.
16. Goodman, R. (2001). Psychometric Properties of the Strength and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.
17. Goodman, R., Ford, T., & Meltzer, H. (2002). Mental Health Problems of Children in the Community: 18 Month Follow Up. *BMJ*, 324(7352), 1496-1497.
18. Goodman, R., Simmons H., Gatward, R., & Meltzer, H. (2000). Using Strengths and Difficulties Questionnaire (SDQ) to Screen for Child Psychiatric Disorders in a Community Sample. *The British Journal of Psychiatry*, 177(6), 534-539.
19. Groth-Marnat, G & Wright, A. J. (2024). *Handbook of Psychological Assessment*. Wiley.

20. Hoosen, N., Davids, E. L., de Vries, P. J., & Shung-King, M. (2018). The Strengths and Difficulties Questionnaire in Africa: A Scoping Review of its Application and Validation. *Child and Adolescent Psychiatry and Mental Health*, 12.
21. Hunsley, J., & Mash, E. J. (2007). Evidence Based Assessment . *Annual Review of Clinical Psychology*, 3, 29-51.
22. International Testing Commission. (2023). ITC Guidelines for Translating and Adapting Tests. *International Journal of Testing*, 23(2), 89-134.
23. Jay, M. A., Sanders-Ellis, D., Blackburn, R., Deighton, J., & Gilbert, R. (2023). Umbrella Systematic Reviews Find Limiting Evidence that School Absence Explains the Association between Chronic Health Conditions and Lower Academic Attainment. *Frontiers in Public Health*, 11, 1122769.
24. Kankaanpää R., Totto, P., Punamäki, R.-L., & Peltonen, K. (2023). Is it Time to Revise the SDQ? The Psychometric Evaluation of the Strengths and Difficulties Questionnaire. *Psychological Assessment*, 35(12), 1069-1084.
25. Kankaanpää, R., Töttö, P., Punamäki, R.-L., & Peltonen, K. (2023). Is it Time to Revise the SDQ? The Psychometric Evaluation of The Strengths and Difficulties Questionnaire. *Psychological Assessment*, 35(12), 1069-1084.
26. Karlsson, P., Larm, P., Svensson, J., & Raninen, J. (2022). The Factor Structure of the Strength and Difficulties Questionnaire in a National Sample of Swedish Adolescents: Comparing 3 and 5 factor models. *PLOS ONE*, 17 (3), E0265481.
27. Kashala, E., Elgen, I., Sommerfel, K., & Tylleskar, T. (2005). Teacher Reports of Behavioural and Emotional Problems among Schoolchildren in Kinshasa, Democratic Republic of Congo. *European Child and Adolescent and Adolescent psychiatry* , 14, 381-387.
28. Khawaja N. G., & Dhushyanthakumar. (2020). Strengths and Difficulties questionnaire-Teacher Investigating Factor Structure and Utility with Culturally and Linguistically Diverse Students. *Journal of Psychologists and Counsellors in Schools*, 30(2), 199-215.
29. Liegmann, K., Fischer, L., Dadaczynski, K., Hanewinkel, R., & Morgenstern, M. (2025). Reliability and Validity of the Self Report Version of the Strength and Difficulties Questionnaire (SDQ) in Primary School Children. *International Journal of Behavioral Development*, 49(4), 406-414.
30. Lord, F. M., Norvick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley.
31. Margaretha, M., Azzopardi, P. S., Fisher, J., & Sawyer, S. M. (2023). School Based Mental Health Promotion: A Global Policy Review. *Child and Adolescent* , 14, 1126767.
32. Mbithi, G., Mabrouk, A., Sarki, A., Odhiambo, R., Namuguzi, M., Dzombo, J. T., Atukwatse, J., Kabue, M., Mwangi, P., & Abubakar, A. (2023). Mental Health and Psychological Well-being of Kenyan Adolescents from Nairobi and the Coast Regions in the Context of COVID-19. *Child and Adolescent Psychiatry and Mental Health*, 17, 63.
33. Mokkink, L. B., Elsmann, E. B., Terwee, C. B. (2024). COSMIN Guidelines for Systematic Reviews of Patient-Reported Outcome Measure Version 2.0. *Quality of Life*, 33(11) 2929-2939.
34. Ortuno-Sierra, J., Fonseca-Pedrero, E., Paini, M., Sastre i Riba, S., & Muniz, J. (2015). Screening Mental Health Problems During Adolescence: Psychometric Properties of the Spanish Version of the Strengths and Difficulties Questionnaire. *Journal of Adolescence*, 38, 49-56.
35. Ortuno-Sierra, J., Fonseca-Pedrero, E., Paino, M., & Sastre i Riba, S. (2022). Cross-Cultural Validation and Measurement Invariance of Strengths and Difficulties Questionnaire: A Systematic Review. *Current Psychology*, 41, 6154-6170.
36. Sawyer, M. G., et al. (2024). Validity of Strengths and Difficulty Questionnaire for Screening and Diagnosis in Western Australian Adolescents. *Diagnostics*, 14(21), 2433.
37. Sheel, H., Suarez, L., & Marsh, N. V. (2025). The Factor Structure of Strengths and Difficulties Questionnaire in Indian. *Assessment*, 43(3).
38. Siva, T. B. F., Osorio, F. L., & Loureiro. (2024). Discriminative Validity and Diagnostic Potential . *Frontiers in Psychology*, 6, 811.
39. Stone, L. L., Otten, R., Engels, R. C. M. F., Vermulst, A. A., & Janssens, J. M. A. M. (2015). Psychometric Properties of the Parent and Teacher Versions of Strength and Difficulties Questionnaire: A Review of the Literature . *Clinical and Family Psychology Review*, 18(3), 254-274.

40. Terwee, C. B., Prinsen, C. A., Chiarotto A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., Mokkink, L. B. (2018). COSMIN Methodology for Evaluating the Content Validity of Patient-Reported Outcome Measures: A Delphi Study. *Quality of Life Research*, 27(5), 1159-1170.
41. Theunissen, M. H. C., de Wolff, M. S., Eekhout, I., Mieloo, C. L., Stone, L. L., & Reijneveld, S. A. (2022). The strengths and Difficulties Questionnaire Parent: Dutch Norms and Validity. *BMC Paediatrics*, 22 Article 202.
42. Van de Vijver, F. J. R., & Tanzer, N. K. (2023). Cross Cultural Assessment- Principles and Applications. In D. H. Saklofske, V. L. Schwane, & C. R. Reynolds (Eds), *The Sage Handbook of Psychological Assessment*. SAGE Publications.
43. Vugteveen, J., De Bildt, A., Timmerman, M. E. (2022). Normative Data for the Self Reported and Parent-Reported Strengths and Difficulties Questionnaire (SDQ) for ages 12-17. *Child and Adolescent Psychiatry and Mental Health*, 16, Article 5.
44. Wilcox, G., Schroeder, M., & Drefs, M. A. (2023). Clinical Reasoning: A Missing Piece for Improving Evidence Based Assessment in Psychology. *Journal of Intelligence*, 11(2), 26.
45. World Health Organization. (2024). Adolescent Mental Health. WHO.
46. Wright, A. J., Pade, H., Gottfried, E. D., Arbisi, P. A., McCord, D. M., & Wygant, D. B. (2022). Evidence Bases Clinical Psychological Assessment (EBPA): Review of Current State of Literature and Best Practices. *Professional Psychology: Research and Practice*, 53(4), 372-386.
47. Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2023). Evidence-Based Assessment in Child and Adolescent Mental Health: Recent Advances and Future Directions. *Journal of Clinical and Child Adolescent Psychology*, 52(5), 689-704.
48. Zhong, C., Huang, J., Luo, Z., Lian, C., Zhou, M., & Kuang, L. (2022). Development and Validation of Rapid Assessment Survey of Primary Care in China. *Frontiers in Public Health*, 10, 852730.