

Review and recommendations for national
policy for England for the use of mental
health outcome measures with children and
young people

**Report for:
Department of Children Schools and Families
and Department of Health**

Tender No: 2008/056

**Final report
31st October 2008**



Institute of Psychiatry
at the Maudsley

MANCHESTER
1824



Contact:

Dr Miranda Wolpert
Director Child and Adolescent Mental Health Services (CAMHS)
Evidence Based Practice Unit (EBPU)
University College London and the Anna Freud Centre
21 Maresfield Gardens, London NW3 5SU
Tel: +44 (0)20 7443 2218, e-mail: mirandawolpert@camhs.org.uk

Research group

Dr Miranda Wolpert, EBPU, University College London and the Anna Freud Centre

Professor Peter Fonagy, University College London and the Anna Freud Centre

Professor Norah Frederickson, University College London

Dr Crispin Day, Institute of Psychiatry, King's College London

Professor Sir Michael Rutter, Institute of Psychiatry King's College London

Dr Neil Humphrey, University of Manchester

Professor Panos Vostanis, University of Leicester

Dr Pam Meadows, National Institute of Economic and Social Research

Dr Tim Croudace, University of Cambridge

Professor Peter Tymms, Durham University

External consultant: Dr Jeb Brown, Center for Clinical Informatics, Salt Lake City, Utah

Systematic Reviewers:

Dr Jessica Aitken, EBPU, University College London and the Anna Freud Centre (lead)

Hayley Syrad, EBPU, University College London and the Anna Freud Centre

Maria Munroe, EBPU, University College London and the Anna Freud Centre

Cathy Saddington, EBPU, University College London and the Anna Freud Centre

Emma Trustam, CORC/EBPU, University College London and the Anna Freud Centre

Jenna Bradley, CORC/EBPU, University College London and the Anna Freud Centre

Dr Sevasti-Melissa Nolas, EBPU, University College London and the Anna Freud

Centre

Paula Lavis, EBPU, University College London and the Anna Freud Centre

Dr Alice Jones, University College London

Contents

	Page
Researchers	1
Section 1: Executive Summary	3 - 5
Section 2: Introduction: Overview of Key issues	6 - 13
Section 3: Review Method Overview	14
Section 4: Review of existing measures	15 - 25
Section 5: Consultation with Stakeholders	26 - 27
Section 6: Review of outcome measures for performance management	28 - 31
Section 7: Consideration of Implementation issues in England	32 - 38
Section 8: Conclusions and Recommendations	39 - 43
Section 9: Table summarising proposed way forward	44
Section 10: References	45 - 47
Appendix 1: Lists of measures identified from initial search of databases and stakeholder consultation	48 - 49
Appendix 2: Quality assurance processes for review of existing measures	50
Appendix 3: Comments from stakeholders in response to consultation	51- 55
Appendix 4: Comments from CORC members in response to consultation	56 - 60
Appendix 5: Potential conflicts of interest	61

Executive Summary

Remit

To carry out a review of measures in the area of children's psychological well-being and children's mental health, in order to advise the Departments of Children Schools and Families (DCSF) and Health (DH) on which are the best measures that could be implemented from April 2009 onwards, (as part of Public Service Agreements and other means) to influence the system to promote better psychological well-being and effective intervention for those children with mental health problems, including, if appropriate, advising on how to take forward the development of a new approach or measure.¹

Method

Four strands of work were undertaken (June- October 08):

- a) Systematic search and structured review of existing outcomes measures in child mental health
- b) Consultation with stakeholders about their experiences of using measures
- c) Review of use of outcome measures for driving up performance
- d) Consideration of implementation issues

Key learning

Learning from the systematic search and literature based review of measures:

- a) It was possible to identify a small number of psychometrically robust measures that could potentially be considered for use alongside other measures and data collection, to influence the system to promote better psychological well-being and effective intervention for those children with mental health problems.
- b) However, whilst the number of measures that are available to assess child mental health outcomes is growing fast (and there are some interesting new measures in development), all existing measures identified had limitations as well as strengths and none met the full range of psychometric criteria likely to be essential for successful ensuring they can reliably measure both severity and change over time in key groups. In particular, no measure has been tested for bias or differential performance in different UK ethnic minorities and other hard to reach groups and we found no evidence of any single measure being successfully used for performance management of the kind foreseen by this initiative.
- c) It can be difficult to adapt any measure for which the copyright is held by one academic, institution or psychometric corporation to meet a local or particular need, and this potentially puts the academic or institution in a position to "hold hostage" anyone who wants to use and adapt such a measure.
- d) Looking at all existing measures there is a considerable degree of conceptual overlap in terms of item wordings (there may be only so many ways you can ask if someone is feeling sad).

¹ In August 2008 the brief was extended to include a consideration of measures of social and emotional skills of children and young people. This aspect of the report is to be completed by December 2008.

- e) Given the latest psychometric techniques and the level of knowledge now available one possible way forward may be to construct an item bank that can be used to assess different aspects of mental health in children more flexibly in the future without them all necessarily being given all the same items.

Learning from stakeholder consultation:

- a) There is evidence that some of the key measures are now in widespread use and becoming embedded in or alongside systems that can monitor and feedback on outcomes achieved and practitioners report on them positively though recognising their limitations.
- b) It is important to let stakeholders have a degree of choice to use additional complementary measures for their specific populations or contexts.
- c) It is important in the long term to develop systems that allow for flexibility of assessment portfolios and possibility for change over time.

Learning from the review of use of measures to drive up performance:

- a) In order to mitigate against the possibility that any implementation strategy is open to perverse incentives, particular attention needs to be paid to how information is collected (sampling bias), entered (data accuracy), analysed (recognising the need for realistic complexity) and interpreted (the validity of inferences).
- b) It is vital to analyse and try to make sense of data from any measure or assessment process in the light of appropriate contextualising information (much of which is already available) and multiple sources of information.
- c) It is crucial to build systems that allow for integration of data with other sources and for triangulation of results and findings.
- d) It is dangerous to rely on any single measure as a stand alone assessment as this can all too easily lead to perverse effects and make it difficult to interpret data.

Learning from our review of implementation issues:

- a) Online may be the best and most cost-effective mode of implementation wherever possible.
- b) National dataset development IT developments in England, including the development of electronic care records, means that some currently used measures are now being built into future national systems as well as already being built into a range of existing local and regional systems.
- c) Commissioners should allow a minimum of 3-5% of budget spend to ensure outcome evaluation can occur in services.

Key Recommendations

1) Population survey

To gauge population level of well being and access to support consideration should be given to the use of a biannual population survey of children aged 8 and above in schools and their parent. This could involve using a quality of life measure such as

Kidscreen alongside additional questions about **access to help** and possibly a measure of **social and emotional skills**.²

2) Service level outcomes evaluation

To evaluate service level outcomes in order to help assess the effectiveness and drive up performance, routine evaluation should be mandated in a way that ensures services do collect common outcome data but that also allows flexibility. Starting with a focus on specialist Child and Adolescent Mental Health Services (CAMHS) for 3-17 year olds; the **Strengths and Difficulties Questionnaire (SDQ)** (parent and child versions) should be used **with two or more additional measures to be decided by services as relevant to their specific settings** such as: goals based measures, measures specific to the presenting difficulties, measures of practitioner views (such as the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) or Child Global Assessment Scales (CGAS) and measures of outcome of consultation to others as relevant. Data from these should be considered alongside other relevant data including measures of experience of services (such as Commission for Health Improvement, Experience of Service Questionnaire (CHI-ESQ)). Commissioners should allow 3-5% of budgets for implementation of outcome evaluation across services. Universal and targeted services should be encouraged to use outcome evaluation with a suggestion this include goals based outcome measures as a minimum.

3) Future development of a new measure/approach

To develop a new assessment method that will allow greater coverage and flexibility of future adaptation a **tender for the development of a new measure/approach** is suggested. The tender should specify copyright and intellectual property to rest with DCSF/DH.

4) Establishment of a child well being observatory/group

To develop a means to make use of nation-wide data collection on psychological well being and mental health problems to support valid inferences from the data through appropriate analysis and future research and development, consideration should be given to the **establishment of a child well being observatory or group**. This body or group should include psychometric and child mental health specialists who can aid DCSF/DH in examining data and making sense of it in a meaningful way advise on future developments and innovations.

5) Communicating with the field

In communicating with the field the DCSF/DH should stress the complexities of the issues and the limitations of any subjective measures, and emphasise the need to be cautious about any interpretation of outcome data from a single source and the importance of viewing it in context of other relevant data.

The details, timescales and implementation issues that need to be taken into account in relation to these recommendations are summarised on **pp 41- 43** below.

² to arise out of review of measures of social and emotional skills to be completed December 2008

Section 1: Introduction: Overview of key issues

Given that the meaning of “outcome measures” for child³ mental health could be interpreted in a number of ways, it is considered helpful to spell out some of the different possible ways forward at the outset in relation to five key overarching issues:

- 1) The settings in which the measure might be used
- 2) The scope of the measure in terms of the aspects of mental health it covers
- 3) Who provides the information relating to the child’s mental health
- 4) The implications of different forms of administration
- 5) The different purposes for which the information may be used

1) Setting

Measurements of mental health outcomes have been developed to access a variety of different populations. On the broadest level, the population of interest has been the **general population**. These populations, found in universal settings, incorporate data from children irrespective of risk factors or clinical diagnoses and typically contribute to information on epidemiology/prevalence rates. Access to these populations is often gained through environments the child accesses independent of any mental health issues such as the school environment and other community settings. Measurement of mental health outcomes in these settings typically relies on information from the child themselves with the possible addition of parent reports, and/or teacher reports where schools are the point of access.

Other populations of interest include more **targeted populations** of those children at risk of mental health problems. These groups may be targeted by virtue of several features, including membership of particular demographic groups identified as at risk of specific mental health problems (e.g. children of parents with mental health problems, children looked after by the state), or by sampling children who have been previously identified as experiencing some adjustment problems (e.g. perhaps those identified as behaviourally disruptive in classrooms). Who provides responses in this setting is largely dependent on the method of identification of the target group but options include parent, child and teacher reports with the addition of other workers who may be in contact with the target groups (e.g. youth workers, social workers).

Finally, information may be required regarding children accessing **specialist mental health services**, or those who have diagnosed mental health difficulties. Therefore, measures of mental health for this population are often administered within a clinical setting. Information relating to these groups can be provided by children, specialist mental health practitioners and parents. Instead of providing indications of prevalence, this information is more likely to be used to assess impact of services or changes in levels of adjustment problems before, during and after intervention.

It is highly likely that no one measure will be suitable for all these settings and the research team will aim to advise whether different measures may be necessary for different purposes.

³ Child will be used throughout to refer to children and adolescents.

2) Scope of measure

Given the broad range of settings and purposes for mental health outcome measures, the information required may vary considerably in terms of its breadth and its specificity. For instance, investigations of population-wide prevalence may require a broad measure of mental health in order to capture the wide range of adjustment indices that exist, whereas investigations of service/intervention impact may require a more specific tool to reflect the specificity of the programme being assessed.

Broad indicators of mental health typically provide a general indicator of psychological distress by considering a constellation of behavioural and emotional indicators. These broad measures may have more specific subscales but have more commonly been used as more general screening measures (e.g. Strengths and Difficulties Questionnaire, SDQ), giving an indication of the level of risk for mental health problems. They also tend to be quite brief and, therefore, are generally easy to administer. Such measures can provide a useful 'barometer' in relation to general mental health on a population or individual level. However, they may be less effective in assessing change in symptoms where specific psychological difficulties are concerned because the number of questions tapping into particular symptoms profiles may be insufficient.

More **specific** instruments give a detailed account of particular profiles of mental health problems (e.g. anxiety, depression). They usually provide a more comprehensive indication of a child's functioning in relation to the particular symptoms that together define a particular problem presentation or diagnosis. Such instruments only enquire about a narrow range of symptoms rather than covering a wide range of potential adjustment difficulties (or strengths e.g. Children's Depression Inventory, Kovacs, Spence Anxiety Scale). These measures may be more appropriate for targeted settings where children are either at risk of or have been diagnosed with specific mental health problems. There is some evidence that more specific measures may be better able to track change for specific problems (Lee, Jones, Goodman & Heyman, 2005).

Some measures do provide features of both broad and specific indicators by having specific subscales, which can be aggregated to give a global index of mental health (e.g. Achenbach System of Empirically based assessment- Youth Self Report Scale (ASEBA- YSR)). These measures are, therefore, more flexible in terms of use in different settings and information derived. However, they are typically lengthy due to the combination of breadth and specificity, which may make them more time-consuming and, as a result, less acceptable and potentially more burdensome to users.

A possible approach to addressing the competing requirements of breadth vs. specificity in order to provide a flexible tool for outcome measurement may be partly through the use of a bank of items. **Item banks** contain a large number of items at varying levels of severity on a dimension from which individual items or questions can be selected to create customised shorter scales (Cella, Gershon, Lai, & Choi, 2007). Such flexibility of instrument design and construction might enable one to retain a core set of items collected across settings, which provide a "severity barometer" of child mental health issues as well as defining some more specific outcomes with "optional" items tapping into a range of factors including more specific adjustment problems, positive aspects of child well-being, broader factors such as quality of life and socio-emotional skills.

3) Who provides information?

As alluded to above, the respondent providing information relating to child mental health may partially be determined by the settings within which the measure may be used. The respondent also has a bearing on the costs of administration and the simplicity of language required.

Specialist mental health practitioner reports can be beneficial because these respondents are well placed to provide detailed information on child mental health; they are also better equipped to interpret scores derived from outcome measures. However, specialist mental health practitioner reports are largely currently limited to help-seeking populations in clinical or other specialist settings and ratings often based on reported rather than observed symptoms or difficulties. Use of clinician/practitioner time for information on outcomes measures for more general populations would most likely be prohibitively expensive.

Teacher reports have traditionally been used for research relating to general population mental health, where the school setting has been the point of access to the population of interest. Research suggests that teachers are accurate reporters of children's externalising symptoms (e.g. aggression, conduct disorder); however, they are less well able to provide accurate information on children's internalising symptoms (e.g. depression anxiety), perhaps due to the differential salience of these two indices of adjustment within the classroom (Atzaba-Poria, Pike & Barrett, 2004; Gardiner, 1994; Stanger & Lewis, 1993).

Parent reports have also been employed routinely in mental health outcomes evaluation. They have advantages because they can be accessed irrespective of the setting and are often relied upon when children are considered too young to provide self-reports (e.g. Levitt, Saka, Romanelli, & Hoagwood, 2007). However, there may be some possibility of bias due to parents own mental health status (Conrah, Sonuga-Barke, Stevenson & Thompson, 2003) and parent's lack of awareness of internalising difficulties (Verhulst & Van der Ende, 2008). There may be particular difficulties recruiting and retaining parent respondents, particularly from some families where there are complex mental health issues (Littell et al, 2005).

There are strong arguments for the use of **child self-report** as a key perspective. Recent UK policy and legislation has placed increasing emphasis on the importance of the child's perspective across the full range of situations and conditions (e.g. Every Child Matters, DfES, 2004; Children Act, 2004) and the importance of the contribution of children's own views to understanding child mental health problems and what might constitute successful strategies to alleviate these has been stressed (Raby 2007). It has been argued that children are the most practical source of data from universal settings where more general populations are concerned (Levitt et al., 2007). The possibility of eliciting child self report has also been extended by recent developments in terms of 1) research concerning the age at which a child develops accurate self-perceptions and 2) the development of online questionnaires with sound available making the administration of child-self report measures with younger age groups a more viable option (Merrell & Tymms, 2007).

However, there are limitations to the use of child self-reports of psychological

adjustment. In particular: 1) younger children may be more likely to give socially desirable responses about their own mental health than other reporters may be; 2) children with a range of behavioural and emotional problems may be less self aware of these than others around them; 3) young children are less likely than other reporters to be able to read text-based self-report measures or to understand the language or the concepts used in self-report measures; 4) younger children are reportedly less consistent in their self-perception in relation to mental health difficulties and typically respond based on 'the here and now' rather than based on relatively stable levels of psychological adjustment (Roy, Veenstra & Clench-Aas, 2008).

One additional way forward is **child-other report or peer review**. This approach involves children rating other children in their class in relation to aspects of psychological adjustment, although peer ratings of interpersonal relations are also possible (Card et al, 2005). Children as young as 8 or 9 have been shown to be more accurate relative to behaviour observation than either parent or teacher report. However, schools sometimes raise ethical concerns about asking children to rate their peers, and while studies using several different methods have failed to find any evidence of negative repercussions (Bell-Dolan, Foster & Silora, 1989; Mayeau, Underwood, & Risser, 2007), such concerns are likely to make widespread use of this approach untenable in the UK currently.

Best practice suggests a variety of views should be collected and that it is likely to be important to collect more than one viewpoint. It should be noted however, that perhaps because of the different limitations of each measurement approach there is low concurrence between different perspectives (Verhulst & Van der Ende, 2008).

4) Different forms of administration

Both the setting and the respondent have a bearing on the potential cost of administering measures. Mode of administration is also an important determining factor. Typical administration methods include postal, telephone, face-to-face and, more recently, online completion. Each of these carries different cost implications.

Online survey facilities typically cost between £500 and £1,000 a year but the price varies based on the degree of personalisation required. Additionally, the set-up (including checking and testing questionnaires using existing instruments) amounts to approximately £500. However, there are usually no costs in addition to the subscription; respondents complete data online and a dataset is created automatically. As the use of online surveys has only become widespread in the last few years, academic articles discussing response rates remain relatively rare. Early articles (around 2001) suggested response rates were lower than for postal methods (see below). These were typically based on college students. More recent comparisons find that response rates are similar between the two methods, including surveys involving children and young people (Truell, Bartlett, & Alexander, 2004; McCabe, Boyd, Young, Crawford, & Pope, 2005).

According to the latest Ofcom survey (http://www.ofcom.org.uk/advice/media_literacy/medlitpub/medlitpubrss/ml_childrens08/cannex.pdf) 71% of 8-11 year olds and 77% of 12-15 year olds live in homes where there is access to the internet. Of these more than 90% have broadband access. In

addition to this, some parents will also have access at work and most people will have access via a local public library or UK Online centre, or via friends or family members. Children and young people will generally have internet access at school. This means that an online completion system would have the potential to reach most respondents at low cost.

Postal surveys generally cost at least 80p - £1 per head for printing, envelopes and postage (including staff time for putting surveys in envelopes). However, some postal surveys have been estimated at twice this amount. For postal surveys, 40% to 50% would represent a good response rate, though between 25% and 30% would be more common. Furthermore, it is essential to include a FREEPOST envelope for the response, and evidence suggests that stamps yield a better return rate than franking on the outward mailing (Edwards, Roberts, Clarke, DiGuseppi, Pratap, Wentz, & Kwan, 2002). With this form of administration, the real expense lies in manually coding and entering data once returned, which would be partly determined by the length of the instrument used and the number of pre-coded versus open-ended questions. For example, for a 50-item instrument where all answers were pre-coded data entry costs would be approximately £5 for each. Coding and open-ended questions would add to this, and could potentially double the cost, resulting in an average cost of around £10 per completed survey would be typical.

Telephone surveys accrue costs in terms of interviewers, but can be set up to automate the data capture via Computer Assisted Telephone Interviewing (CATI). Programming costs for this kind of approach should allow approximately £1,000 for programming plus £10 per completed survey. This automation has meant that the cost differential between postal and telephone surveys has all but disappeared. However, a lot of time is often spent locating respondents and setting up appointments to talk. Survey companies tend to be able to use time effectively by running several surveys simultaneously, and also by having large numbers of respondents so that down-time for interviewers is minimised. Response rates are generally slightly below those for face-to-face surveys (around 70 per cent would be typical) (de Leeuw & van der Zouwen, 1988).

In terms of **face – to - face completion**, assuming that is achieved via Computer Assisted Personal Interview (CAPI) so that data capture is automated, programming cost would be the same as with CATI above (i.e. around £1,000). Additional completion costs would then depend on how much staff time is involved and the job description of the staff member. For example, typical costs for a 30 minute instrument would be: clinical psychologist, £33.50; mental health nurse, £20.50; mental health social worker, £40.00; generic CAMHS worker, £31.00; generic multi-disciplinary CAMHS team member, £43.00 (Personal Social Services Research Unit (2008) *Unit Costs of Health and Social Care, 2007*); teachers, £12.50 (based on typical supply rates of £25 per hour). A general rule of thumb using survey interviewers rather than professionals would be that a face-to-face survey costs roughly twice what a telephone survey costs (Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau, 2004). Response rates are generally around 75 per cent (de Leeuw & van der Zouwen, 1988).

It is worth bearing in mind that if an instrument forms part of a practitioner's assessment of a child or young person's needs, then the completion of the instrument forms part of the treatment package and is not additional to it. However, the length of the instrument

and the requirements in the case of certain instruments for particular qualifications in those who complete them means that the costs of clinical time will be taken into account when it comes to considering the costs and benefits of different instruments.

All methods have advantages and disadvantages. Face-to-face and telephone surveys are likely to be better at capturing those with poor literacy. If available in different languages they can also be more accessible to those with poor English. Online surveys will tend to under-represent those who do not have internet access at home, since the effort required to complete the survey is likely to be greater for this group. Telephone surveys generally rely on the use of land lines, and one-third of low-income households below the age of 65 do not have a land line, although it is likely that this figure is lower among households with children (Ofcom, 2007).

Whatever method is used, response rates are better if the respondent is interested in the issues, the survey is easy to understand, does not take too long to complete, and the approach to the respondent is personalised. In general, respondents tend to prefer methods that do not involve interviewer contact (i.e. postal and web-based) for sensitive subjects (Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau, 2004).

From the descriptions above it is clear that costs extend greatly beyond simple purchase of the measure to be completed. Therefore, any cost analysis needs to factor in potential set-up costs, running costs, support staff and data entry/management. These will be considered in detail in the current report.

5) Purposes for which the data may be used

The suitability of different measures depends on what the data are to be used for.

If the aim is to tap **psychological well being in a community** then positive measures of well being may be the most relevant approach. Measures such as Kidscreen have been developed to give a broader sense of child well-being, taking into account contextual factors including consideration of home and school settings, interactions with peers and other adults and general health. These measures are typically created to capture a more general quality of life (Eid & Diener, 2004; Eid, Schneider, & Schwenkmezger, 1999) than that traditionally offered by more focused mental health instruments which emphasise symptoms and poor psychosocial functioning and therefore may not capture important variations in aspects of well being experience by psychologically well individuals. However, such measures may have weaker psychometric properties than existing measures of problems or symptoms and may be less likely to show change in response to specific interventions.

If the aim is to drive up **individual practitioner performance** then getting feedback on progress for individual children and families over time as compared to others seeing similar clients has been suggested as a powerful way forward (Miller, Duncan & Hubble 2004, Lambert & Ogle 2004). However, the psychometric strengths of this approach are still open to question and existing outcome measures may not be sufficiently psychometrically robust to inform individual patient care and care needs to be taken to ensure appropriate caseload assessment and benchmarking procedures are put in place Minami and Brown (2008).

If the aim is to drive up **service performance** of agencies and services then it may be that measures most closely related to change in the population receiving the service may be the best approach. The basis of this type of measure is to set goals and measure the progress towards achieving these goals. Measures of this kind are accessible to and appropriate for all children/young people irrespective of their age or presenting problem(s). The setting of goals allows progress to be measured which is specific and relevant to the child/young person, and articulating these aims can help focus the treatment/intervention on achieving the desired end result, which may otherwise remain implicit. It has been argued that goal attainment measures increase involvement and in doing so increase commitment. However, in setting individual goals, a goal attainment measure will not be standardized in the way that outcomes measures are, making the issue of comparing results across individuals and looking at the robustness of the measure more problematic in psychometric terms.

A particular factor to take into account in considering the suitability of measures for use in service improvement contexts is their ability to be used with appropriate **risk adjustment or value added algorithms**. The importance of risk adjustment in health care has been stressed by lezzoni (2003) and others. It is argued that the meaningful comparison of patient outcome requires adjustment for those patients' risk factors. Risk adjustment "levels the playing field," accounting for factors that patients bring to the healthcare encounters that could affect their outcomes. Controlling for patients' risk factors is essential for drawing useful inferences from observed outcomes of any intervention in relation to the treatment effectiveness, provider performance, or quality of care (lezzoni, 2003). In the field of child mental health it is not yet clear what the relevant "risk factors" are, since issues such as case complexity or severity of difficulties are not always associated with outcomes in the ways predicted but the general principle of being able to allow for relevant mediating and moderating factors in interpreting outcome data are as relevant in this area as others.

Finally, the literature suggests that if outcome data are to be used to drive up performance then practitioners have to be able to see the benefits of this approach. An earlier review undertaken in relation to the mental health outcomes initiative for adults (Fonagy, Matthews & Pilling, 2004) concluded that before routine outcomes measurement is likely to be successfully implemented the core benefits accruing to individual clinicians and teams in relation to clinical management would need to be established. The report also recommended that minimum standards and requirements would need to be set nationally together with providing the supporting framework to encourage services to go beyond these minimum standards and develop innovative local use of outcomes data.

Implications for the review of these overarching issues

In the light of these issues the following decisions were taken at the outset by the research team following consultation with colleagues in DCSF/DH:

1) Setting

To look for a measure or measures that can be used from universal to specialist settings. However, to be clear if this not possible about possible use of different measures in different settings.

2) Scope

To look first for a broad measure and only revert to looking for more specific ones if a suitable broad measure could not be found. Moreover, to look for a measure that had some positive elements along with strong psychometric properties.

3) Who provides the information?

To look for a measure that could either be completed by parents or children and/or by practitioners as part of their routine work. Measures that could only be completed by specialist practitioners or teachers where this was in addition to their routine work were excluded on the grounds of cost.

4) Different forms of administration

To look for measures that could be completed online, on paper or over the telephone. To exclude measures that could only be completed face to face as part of an interview on the grounds of cost.

5) Use of information

To look for measures that could be used to access a variety of different forms of information and to pay particular attention when reviewing current uses of measures for nationally directed service improvement and performance management and that will provide meaningful information at a local level to practitioners and other stakeholders.

Section 2: Review Method Overview

The research team undertook four strands of work

1. Systematic search and structured review of the published literature of outcomes measures in child mental health including reviewing identified measures in terms of “psychometric properties” and “implementability”
2. Consultation with stakeholders about their experience of measures
3. Review of use of outcome measures for driving up performance
4. Consideration of implementation issues in England

The method, findings and conclusions from these four strands of work are reported below.

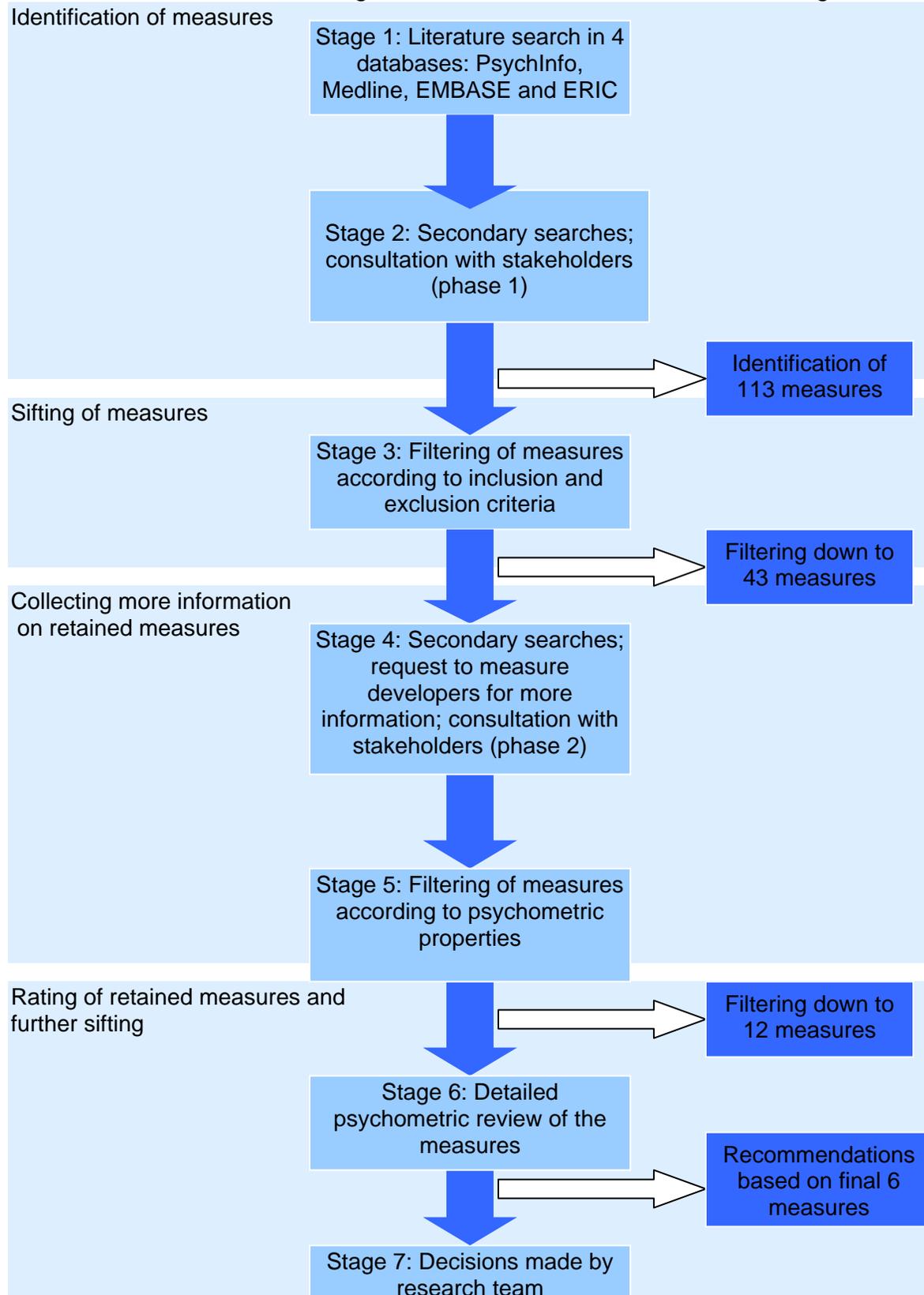
Note on Methodological limitations

The tight timescale for this review imposed inevitable constraints. The commission from the DCSF/DH was for the research team to make specific recommendations based on their best understanding of the information that they had been able to review and consider within the timescale. Although the search strategy was entirely systematic, we were only able to obtain a subset of all available materials identified by the review and to consult with stakeholders through existing networks. The findings should therefore be viewed as based on a rapid and critical appraisal of the information that was available to us at the time. It would have taken considerably longer to systematically review and consult on all the available information on the instruments/measures identified. Such activities are rarely attempted for a group of instruments at one time, usually for a single measure (not the several dozen identified here).

Section 3: Systematic search and structured review

Method

Process of the review of existing measures is summarised in the flow diagram



Stage 1

Setting parameters of the review

Child mental health outcome measures for the purposes of this review were defined as any questionnaire, measure or approach to measurement that seeks to provide measurement of mental health in children and young people (up to 18 years). Search terms were developed to capture these defining features by a) splitting the terms child mental health outcomes measure into three defining features: I) 'measurement', II) 'mental health' and III) 'child' and b) generating a list of words or phrases that reflect each of these features (see table A below).

Table A:

Factor	Related Terms
1) Measures and approaches to measurement	Measure; questionnaire; survey; checklist; check list; tool; rating scale; scale; repository
2) Mental health and psychological well-being	Mental health; quality of life; psychological adjustment; behaviour problems; emotional problems; mental illness; mental disorder; psychiatric disorder; behavioural and emotional difficulties; social difficulties; social and behavioural difficulties; conduct problems; internalising; externalising; depressive symptoms; antisocial; self-esteem; pride; prosocial behaviour; sense of belonging; hopefulness; well being; positive self-regard; aggression; anxiety; depression; mood; feeling
3) Children	Children; adolescents; paediatrics

Search of key databases

Initial searches focused on 4 key databases: MEDLINE, EMBASE, PsychInfo, ERIC by conducting searches in each database for the terms relating to 1) measures and approaches to measurement, 2) mental health and psychological well-being and 3) children.

Searches for these terms were then conducted in the thesaurus facility within each of the four databases. These thesaurus searches were carried out to identify subject heading or MeSH terms under which papers of interest would be catalogued in each database. Searches in thesauruses of each database identify terms that map onto MeSH headings or subject headings relevant to child mental health outcome measures which are specific to each database. Subject headings and MeSH terms for each database were then refined to discard terms that were too broad or too narrow to capture our search criteria. These final MeSH headings/subject headings were used to run searches in each of the respective databases in order to identify relevant papers (see Table B below).

Table B:

Database	Search terms for subject heading and mesh headings	Initial Hits*	Hits* After Basic Filtering
EMBASE	(Child or children or adolescent) in SU and (child psychiatry or child psychology or mental disease or adjustment disorder or behaviour disorder or emotional disorder or mental instability or mood disorder) in SU)and((psychological test or named inventories or questionnaires or rating scales) in SU) and (LA:EMBV = ENGLISH)	413	90
ERIC	(Check-lists OR Measurement OR Questionnaires OR Rating-Scales).DE. AND (Mental-health OR Behaviour-problems OR Emotional-Adjustment OR Emotional-Disturbances OR Emotional-Problems OR Mental-Disorders OR Psychopathology).DE. AND (Children OR Adolescents).DE. AND LG=ENGLISH	342	58
Medline	((mental health or Mental Disorders or emotional problems or Emotional Disturbances or Behaviour Disorders or adaptation psychological or Psychological adjustment) in MJME)and((child or adolescent) in MJME)and((Questionnaires or psychiatric rating scales or treatment outcome) in MJME)	260	122
Psych Info	(Questionnaires or measurement or surveys or rating scales or Lickert scales or symptom checklists) in SU)and((child psychopathology or child psychiatry or child psychology or adolescent psychopathology or adolescent psychiatry or adolescent psychology) in SU)and((Mental health or community mental health or well being or emotional adjustment or social adjustment or mental disorders or adjustment disorders or psychological stress or behaviour problems or internalization or externalization or psychiatric symptoms or distress) in SU)	96	N/A

*NB 'hits' refers to the number of papers identified in each database

Searches that resulted in over 100 hits were first subject to basic filtering to discard obviously irrelevant hits. This involved the following exclusion criteria being applied to the title:

- that the paper was not related to children's mental health outcome measures
- that the paper was not in English

After this process the titles and abstracts of all remaining hits (or papers) were transferred to a spreadsheet for further sorting based on more specific search criteria. This involved applying the above criteria to any databases returning less than 100 hits and filtering all hits based on further criteria (see Box A).

Box A

Hit excluded if:

- 1) no child mental health outcome measure was mentioned in abstract
- 2) the measure mentioned was too narrow to provide a broad assessment of mental health (e.g. focused exclusively on just personality disorders or just schizophrenia)
- 3) the paper referred to a measure not used with children
- 4) the paper was not in English
- 5) the paper was a duplicate of a previous hit within the database
- 6) the paper referred to an assessment or DSM diagnosis

Hits remaining after this filtering procedure were sorted according to the measures they referred to in order to create a list of measures as per the search strategy.

Stage 2

Secondary searches and consultation with stakeholders

Secondary searches were conducted in order to gain further information. This involved reading through additional abstracts and papers, which discussed the measures and gaining information from the companies marketing the measures and/or measure developers (e.g. information on the focus, age range, reporters etc.).

Consultation with practitioners to identify additional measures and rate identified measures was conducted a questionnaire at a national conference and via e-mail through stakeholder networks. Details of the consultation are given in Section 5 below.

This process resulted in 113 measures being identified (listed in Appendix 1).

Stage 3

Inclusion and exclusion criteria for measures or approaches to measurement

Inclusion criteria:

To include any questionnaire, measure or approach to outcome evaluation

1. that seeks to provide measurement of generic mental health in children and young people (up to 18)
2. that is either multi-dimensional or uni-dimensional
3. that can be completed by child or parents/carers with the possible addition of professionals
4. that has been validated in a child or adolescent context, even if not originally developed for this purpose
5. that is available in English language
6. that can be used with a reasonably wide age range (e.g. not just for preschoolers)

Exclusion criteria:

To include questionnaires, measures or approaches to outcome evaluation

1. that are not available in English
2. that do not measure mental health outcomes
3. that do not cover broad range of difficulties i.e. concern only specific mental disorders or domains e.g. ADHD, schizophrenia, physical problems, eating disorders, self-harm, OCD, psychosis, autism, specific learning difficulties, phobias etc, or internalizing or externalizing only
4. that are not used with children
5. that are based on professional report only (e.g. teacher or clinician);
6. that take over 30 minutes to complete
7. that provide open-ended responses that have to be manually coded
8. where the age range was too narrow (e.g. pre-school version of the measure only)
9. that are not available in English
10. have not been used with a variety of populations (i.e. only used with very specialist groups)

Application of the above criteria resulted in a reduced list of 43 measures. Those measures where there was not enough information yet to judge whether they met the inclusion/exclusion criteria remained included at this stage

Table C Showing 43 measures identified after stage 3

Achenbach System of Empirically Based Assessment (ASEBA) [Child Behaviour Checklist CBCL]
Beck Youth Inventories (BYI)
Behaviour Assessment System for Children (BASC)
Behavioural and Emotional Rating Scale (BERS)
Behavioural and Emotional Strengths Scale (BESS)
Brief Multidimensional Student Life Satisfaction Scale
Child Adaptive Behaviour Inventory (CABI; CBI)
Child behaviour rating form typical IQ version (CBRF-TIQ)
Child Health Questionnaire
Child Outcome Rating Scale (CORS)
Child Symptom Inventories (CSI),
Conners Rating Scales
Index of Child and Youth Well-Being (CWI)
Devereux Scales of Mental Disorders
Dominic interactive
"Feelings Attitudes and Behaviours Scale for Children (FAB-C),
Generic Children's Quality of Life Measure (GCQ).
the MacArthur Health and Behaviour Questionnaire (HBQ)
Health of the National Outcome Scale for Children and Adolescents (HoNOSCA)
KINDL
Massachusetts Youth Screening Instrument- 2
Millon Adolescent Clinical Inventory (MACI)

Missouri Children's Behaviour Checklist (MCBC)
Ontario Child Health Scale (OCHS)
Parent-Child Rating Scale (P-CRS) and Teacher-Child Rating Scale (T-CRS)
Pathology summary score (PATH)
Peabody Treatment Progress Battery (PTPB)
Pediatric Symptom Checklist (PSC)
PedsQL Present Functioning (PedsQL)
"Pictured Child's Quality of Life Self Questionnaire (AUQUEI)
Piers-Harris 2
Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q)
Quality of Life Profile Adolescent Version (QOLPAV)
Reynolds Adolescent Adjustment Screening Inventory
Rutter Scales (Revised Rutter Scales)
School Children's Happiness Inventory (SCHI)
Strengths and Difficulties Questionnaire (SDQ)
Symptom Check List (SCL)
Target Symptom Rating (TSR)
Youth Outcome Questionnaire
Youth Quality of Life Instrument- Research Version or Surveillance version (YQOL)
UK Kidscreen, parent and child report
Young Persons Clinical Outcomes in Routine Evaluation (YP CORE)

Stage 4

In order to obtain information for this stage of evaluation a consistent strategy was developed for any additional searches required (see Box B).

Box B

When searching for information:

- first try to obtain information from the manual for the measure (where this is available)
- if this is not available, look for the information in review papers
- if these are not available search for published papers, prioritising the most recent (primarily searching for the names of measures in titles and abstracts in the 4 databases described above)

If information is particularly hard to track down on certain measures (e.g. those with no published papers or those with a small number of published papers that are hard to obtain)

- request inter-library loans
- e-mail up to 3 authors
- widen the search from the four key databases to other web sources
- measures yielding no return after this should be excluded

In order to ensure that the maximum amount of information about each measure was taken into account, e-mail requests were sent to all measure developers requesting information about the psychometric properties and implementability of each measure (see tables D and E below).

Table D: Information about psychometric properties requested from measure developers

Property	Definition	Reference to evidence of this in published literature or other accessible data (please attach)
Content validity	The extent to which the domain of interest is adequately sampled by the items in the questionnaire and the items are clear and behaviourally specific	
Internal consistency	The extent to which items in the (sub)scale are inter-correlated, thus measuring the same construct	
Criterion validity (comparison with gold standard)	The extent to which scores on a particular questionnaire relate to a gold standard	
Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	
Dimensionality	The extent to which items load on different factors	
Reproducibility	The extent to which the scores on repeated measures are close to each other (absolute measurement error) Test retest reliability	
Sensitivity to change	The ability of a questionnaire to detect clinically important changes over time	
Interpretability	The degree to which one can assign qualitative meaning to quantitative scores beyond that used for construct validity	
Measure had undergone analysis using Item response Theory (IRT) approaches	The extent to which a sophisticated analysis has been carried using modern methods	

In addition information on the non-psychometric features of each measure (e.g. scope, cross-cultural use, and ethics) was requested from measure developers (see table below).

Table E Information relating to implementability requested from measure developers

Dimension	Details	Reference to evidence of this in published literature or other accessible data (please attach)
Settings	Can it be used in universal, targeted and/or specialist settings	
Financial	Cost per copy – if both electronic and paper versions, cite costs for both (if come in packs, number in pack and price per pack)	
Scoring burden	Estimated time to score responses to measure	
Multiple languages	Names of languages available in	
Use with diverse populations	e.g. learning disability, looked after children, ethnic minorities, city vs. rural populations, other	
Age Range	Details of age ranges developed for/used with + evidence	
Respondent versions	e.g. child, parent, teacher, clinician	
Ease of use for poor readers	e.g. reading age required for measure, length/complexity of sentences, response scales	
Length	Estimated time to complete measure and length of measure (number of items)	
Existence of UK norms	Published norms for the UK? Evidence of use in the UK?	

In addition consultation with stakeholders asked about their experience of measures (discussed in detail in Section 5).

Stage 5

After collating all the information from the above processes, the psychometricians in the research team sifted the remaining measures based on the information gained on the quality of the evidence available for the psychometric properties outlined above and taking into account the original inclusion and exclusion criteria.

At this stage the following additional filtering criteria were applied on the advice of the psychometricians, based on the information now available

- 1) Heterogeneity of samples - the measure was only kept in if used in heterogeneous populations of children with mental health issues (i.e. excluded if only evidence of it is on one particular population e.g. only used with children with one specific type of problem or diagnosis).
- 2) Response scales – only included if responses scale is polytomous – not dichotomous - and can be easily coded (simple yes/no checklist or visual analogue scales (VAS) are therefore to be excluded).
- 3) Extent of evidence – only included if measure had more than 5 published studies or if psychometric evidence is available from independent researchers i.e. other than the original developers.

Quality assurance processes for all stages 1-5 of the review are summarized in Appendix 2

After this process, 12 measures were retained that appeared to have satisfactory psychometric properties.

Table F Showing 12 measures identified after stage 5

Achenbach System of Empirically Based Assessment (ASEBA)
Beck Youth Inventories (BYI)
Behaviour Assessment System for Children (BASC)
Behavioural and Emotional Rating Scale (BERS))
Child Health Questionnaire (CHQ)
Child Symptom Inventories (CSI)
Health of the National Outcome Scale for Children and Adolescents (HoNOSCA)
Kidscreen
Pediatric Symptom Checklist (PSC)
PedsQL Present Functioning (PedsQL)
Strengths and Difficulties Questionnaire (SDQ)
Youth Outcome Questionnaire (YOQ)

Stage 6

The remaining 12 measures were passed to two psychometric experts (Tim Croudace and Jeb Brown) to review independently with the full information obtained to date collated on each measure by members of the review team. It was noted that one of the psychometricians had been involved in the development of one of the measures – Jeb Brown in developing the Youth Outcome Questionnaire (YOQ) – though he had never had financial gain from this questionnaire nor was financially or in other ways involved in this measure currently. The two independent reviews of the measures were then anonymised and passed to a third member of the team (Peter Fonagy) who made a summary of their findings.

The review by the psychometricians resulted in six measures being put forward for in-depth consideration by the research group.

Table G Showing 6 measures identified after stage 6

Achenbach System of Empirically Based Assessment (ASEBA)
Child Health Questionnaire (CHQ)
Health of the National Outcome Scale for Children and Adolescents (HoNOSCA)
Kidscreen
Strengths and Difficulties Questionnaire (SDQ)
Youth Outcome Questionnaire (YOQ)

Stage 7

Reaching decisions

Filtering in the previous stages was largely concerned with the identification of measures suitable for measuring child mental health outcomes and screening these according to the robustness of their psychometric properties. This ensured that the measures being considered in the final stages were all of a suitable calibre. Once the most robust measures, for our specific aims, were established final decisions were made on the basis of both psychometric and implementability factors.

In terms of population level evaluation: It was felt the quality of life measures made most sense to use. Both Kidscreen and CHQ had strengths but Kidscreen had more available information and may be more acceptable in terms of range and flexibility of items and possible use in non-health settings, and was therefore put forward as the measure of choice. It was also suggested that a measure of social and emotional skills should be used if possible (to be identified by extended review which reports in December '08) and additional items on service use could be usefully included.

In terms of service level evaluation: Neither Kidscreen or CHQ were regarded as suitable measures for service evaluation in the light of lack of specialist specificity and lack of evidence of sensitivity to change in the way required for use in service evaluation. HoNOSCA was regarded as having strengths as a practitioner report measure but the self report version did not cover a large enough age range or have a strong enough evidence base to be considered as the child report measure of choice. Of the remaining identified relevant three child/parent report measures (SDQ, YOQ and ASEBA) none had been rigorously tested with an extensive range of hard to reach groups nor used as clear performance management tools on their own with proven results. The strong contenders from the United States (YOQ, ASEBA) were rejected because any potential superiority over the UK version (SDQ) was not sufficient to make them worth prioritising and to make up for the fact that they would require reworking systems to incorporate them (see implementation issues), and enough information could not be obtained on the BASC.

In terms of development of a new measure: It was felt that it would be helpful to aim to create a new measure that provides greater coverage of the range of key elements (i.e. mental health, psychological wellbeing and social and emotional skills). DCSF/DH should have intellectual ownership (to allow flexibility of adaptation as relevant) and the measure should be tested with hard to reach groups and possibly used at multiple time

points to be sensitive to change. It should be developed using an item bank model to allow for the greatest degree of flexibility in the future.

Key learning from the systematic search and structured review of measures

- a) It was possible to identify a small number of psychometrically robust measures that could potentially be considered for use alongside other measures and data collection, to influence the system to promote better psychological well-being and effective intervention for those children with mental health problems.
- b) However, whilst the number of measures that are available to assess child mental health outcomes is growing fast (and there are some interesting new measures in development), all existing measures identified had limitations as well as strengths and none met the full range of psychometric criteria likely to be essential for successful ensuring they can reliably measure both severity and change over time in key groups. In particular; no measure has been tested for bias or differential performance in different UK ethnic minorities and other hard to reach groups and we found no evidence of any single measure being successfully used for performance management of the kind foreseen by this initiative.
- c) It can be difficult to adapt any measure for which the copyright is held by one academic, institution or psychometric corporation to meet a local or particular need, and this potentially puts the academic or institution in a position to “hold hostage” anyone who wants to use and adapt such a measure.
- d) Looking at all existing measures there is a considerable degree of conceptual overlap in terms of item wordings (there may be only so many ways you can ask if someone is feeling sad).
- e) Given the latest psychometric techniques and the level of knowledge now available one possible way forward may be to construct an item bank that can be used to assess different aspects of mental health in children more flexibly in the future without them all necessarily being given all the same items.

Section 5: Consultation with stakeholders about their experience of measures

Method

The following phases of consultation were undertaken

Phase 1a

Consultation Questionnaires were given to stakeholders (250) attending the National CAMHS Support Service and Evidence Based Practice Unit Conference (July 08). The questionnaire included a list of measures identified to date and asked respondents to indicate if they had used the measures and if they had found it useful and if there were other measures the research group should be aware of.

- 40 completed questionnaires were returned at the end of the conference July 08

Phase 1b

Further consultation questionnaires were sent out via existing networks of CAMHS stakeholders including the email discussion group FOCUS. This included any additional measures recommended by those who completed the form at the conference and asked respondents to indicate if they had used the measure, if so in what setting and to assess the measures strengths and weaknesses:

- 18 completed questionnaires were returned from email consultation in July 08

Phase 2

A further consultation questionnaire was sent stakeholders in September 08. This listed the reduced list of 12 measures and asked stakeholders about asked respondents to indicate whether they had used the measure for planning provision of services and/or monitoring services and if so how helpful they had found it:

- 11 completed questionnaires were returned from this email consultation in October 08

The list of comments made by stakeholders can be found in Appendix 3.

Findings

Table H below shows numbers of respondents indicating they had used a measure and whether they found it helpful or not.

Table H

Measure	Yes (no. of specialist practitioners finding it helpful)	No (no. of specialist practitioners finding it not helpful)
BERS	1	0
ASEBA [CBCL]	8	1
Conners Rating Scale	23	2
HoNOSCA	10	10
MACI	1	0
SDQ	21	3
YOQ	1	0

Table J below shows average ratings of respondents about how helpful they found the measures in terms of service planning or monitoring?

Table J

Measure	Number who said they had used the measure	Mean helpfulness in service planning (1=not helpful 5= very helpful)	Mean helpfulness in service monitoring (1=not helpful 5= very helpful)
Achenbach System of Empirically Based Assessment (ASEBA)	3	4	4
Beck Youth Inventories (BYI)	11	2.27	2.55
Behaviour Assessment System for Children (BASC)	1	3	3
Behavioural and Emotional Rating Scale (BERS)			
Child Health Questionnaire (CHQ)	1	2	2
Child Symptom Inventories (CSI)			
Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA)	12	2.17	3.25
Kidscreen			
Paediatric Symptom Checklist			
PedsQL Present Functioning (PedsQL)			
Rutter Scales (Revised Rutter Scales)	1	2	2
Strengths and Difficulties Questionnaire (SDQ)	15	2.2	3
Youth Outcome Questionnaire (YOQ)			

The stakeholders responses indicated greatest use of a small number of key measures (SDQ, HoNOSCA, Connors and Beck Youth Inventories rating relatively highly) along with some indication that practitioners felt these were helpful and could be used to help service development and improvement. However, respondents were clearly sensitive to the limitations of different measures and raised the difficulties of meaningful and supported implementation in their comments.

Learning from stakeholder consultation:

- a) There is evidence that some of the key measures are now in widespread use and practitioners report on them positively though recognising their limitations.
- b) Practitioners' comments suggest the importance of having a choice in use of some aspects of measurement to ensure they can use additional specific measure for their population.
- c) It is likely to be important in long term to develop systems that allow for flexibility and change over time.

Section 6: Review of use of outcome measures for driving up performance

Method

A narrative review of use of outcome measures for performance management and service improvement in general and in particular a view of the experience of their use in UK was undertaken with the following findings:

Findings

General principles of performance indicators

The use of indicators for the management of performance is now common across a wide range of public services in Britain. There are a variety of theoretical principles underpinning the selection of appropriate indicators for managing performance, but one of the most widely cited is that of Neely (1998). This suggests that what are needed are:

- Individual measures that quantify the efficiency and effectiveness of actions
- A set of measures that combine to assess the performance of an organisation as a whole
- A system that enables data to be acquired, collated, sorted, analysed, interpreted and disseminated

Smith (2002, 2005) emphasises the importance of ensuring that the measures are designed in such a way that they genuinely indicate whether or not desired objectives are being achieved, and also take account of the structure of incentives within the system. The differing objectives of the various groups involved in a service will require a range of performance management data. For example, patients may use the information to inform their choice of provider. Professionals may require similar data to patients but with the aim of confirming that they are performing according to accepted practice and promoting improvement. Similarly regulators, taxpayers, managers and the Government will have different requirements from performance management measures. This array of needs will therefore necessitate that data content, detail, frequency of collection, presentation and timeliness is tailored to the various stakeholders involved.

The Royal Statistical Society Working Party on Performance Monitoring in the Public Services concluded that performance monitoring done well is broadly productive, but done badly can be harmful (Bird et al, 2005). The working party set out a range of criteria which need to be considered when adopting a measure as a performance indicator. The most relevant criteria in relation to a mental health outcome measure are:

- a) Indicators and definitions should obviate, rather than create, perverse behaviours
- b) Indicators should be straightforward to interpret, avoiding ambiguity about whether the performance being monitored has improved or deteriorated (e.g. because of possible changes in case mix)
- c) Technical properties of the indicator should be adequate e.g. with respect to sampling scheme, response rate in surveys and precision
- d) Indicators should have the statistical potential to exhibit or identify change within the intended timescale of PM [performance monitoring]
- e) Indicators should not impose an undue burden – in terms of cost, personnel

or intrusion – on those providing the information (Bird et al, 2003, pp 6-7).

One fundamental area of concern embodied in criterion (a) above is the implicit assumption that the process of measurement does not influence the behaviour of the individuals and institutions involved in the measurement. This is almost certainly not the case. In economics this principle is known as Goodhart's Law, developed by Charles Goodhart, formerly of the Bank of England, in the context of monetary policy targets. His original formulation stated: "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart, 1975). The point is that in collecting a specific piece of information or observing a particular type of behaviour, the observer or data collector gives information to the person being observed or provider of the information (if only by being present). This knowledge may lead the provider of the information to change their response or behaviour in response to the data collection or observation.

There are growing numbers of examples drawn from healthcare, education, policing and elsewhere of the changes in behaviour induced as a result of measuring a particular indicator (Bird et al, 2005; Pidd, 2005; Propper and Wilson, 2003).

While some distortion of indicators may be unavoidable, one of the key responsibilities of those designing performance monitoring indicators is to recognise that distortion or manipulation might take place, and to ensure that the data collection process has designed out the possibility for manipulation as far as possible (Bird et al, 2005). Ideally, performance measures should be collected independently of the organisation whose performance is being measured (Propper and Wilson, 2003). Performance management theory (see for example, Neely, 1998) would generally suggest the use of multiple indicators in order to capture different dimensions of performance. This reduces the capacity of those being monitored to influence the measures and ensures that outcome measures are looked at in context.

Benchmarking versus performance management

Although the terms benchmarking and performance management are sometimes used interchangeably, they are in fact different. **Performance management** is where an indicator is used by the organisation commissioning or funding a service as an accountability tool for the provider of the service. The indicator is usually published, and failure to meet the specified level of performance may result in reduced funding or even withdrawal of the contract.

Benchmarking, although it may use a similar set of indicators to performance management, is a method of encouraging organisations to improve the quality and quantity of their services by comparing their achievements either with other similar organisations or with the average achievements across a range of providers. The indicators become management tools for use *within* the organisation. They are not a method of providing accountability to a client or funder. Rather they enable an organisation to reflect on its own performance and to consider how any differences from the norm or from best practice can be accounted for.

An indicator may be robust enough to be used for benchmarking, where the incentive for "gaming" is low and where there are genuine gains to be made in professional practice by learning from excellence. But the same indicator may not be as robust when

it is used in the context of hierarchical performance management (Marshall et al, 2000).

Health systems

The use of performance measurement in health systems provides an important means of obtaining improvements and accountability; however their effective implementation also needs to consider a range of related issues. Areas to be taken into account include: the need for further improvement in data collection, analytical methodologies and policy development; the manner by which data presentation influences the interpretation by service users, providers, practitioners and the public; the need for monitoring and counteraction of adverse outcomes of public reporting, for example an avoidance of complicated cases.

Early performance management measures in the health system concentrated on structures and processes. However these data are not useful in the United Kingdom (UK) healthcare policy objective of eliminating unacceptable variations in clinical practice and ensuring uniform high-quality care. The achievement of this requires a change in emphasis away from resources, activity levels and service structures to a focus on outcomes. Clinical outcome measures are the gold standard for measuring effectiveness in healthcare as they are often more meaningful to stakeholders, they ensure that attention and health goals are focused on the patient, and they encourage long-term health promotion strategies (Holloway, 2002; Smith et al, 2008).

Mental health systems

The current use of outcome measurement in mental health systems across countries is variable. A recent Organisation for Economic Co-operation and Development (OECD) technical paper (Garcia et al 2008) describes the results of a survey on mental health information systems conducted in 2007 in eighteen OECD countries⁴. Eleven of the eighteen countries had developed specific mental health information systems, of which eight had some type of system in place and three were planning to implement such a system in the near future.

The systems tended to focus on quality related areas such as outcome/output measurement, benchmarking and clinical guidelines monitoring. Australia, Canada, Denmark, New Zealand, Norway and the USA all have some form of outcome measurement. Australia, Denmark, New Zealand and the USA have benchmarking in place. Denmark, Iceland, the Netherlands and New Zealand have quality/clinical guidelines monitoring in place. Denmark also has an overall staff performance measurement. However, it should be noted that outcome measurement covers a range of results, and more often than not they relate to the measurement of outcome areas such as discharge and suicide rates as opposed to measuring changes in psychological well-being. None of the countries uses a single outcome measure as a performance management tool. Rather, outcome measures are used in conjunction with other measures to monitor system performance.

⁴ The survey was initiated by the OECD Health Care Quality Indicators Mental Health Expert Subgroup which consists of the following 18 fully participating countries: Australia, Canada, Denmark, Finland, France, Iceland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Slovak Republic, South Korea, Spain, Sweden, Switzerland and the United States of America. The United Kingdom, along with the Czech Republic, Hungary and Poland adopted an observer status during 2007.

There are two main reasons for not relying on outcome measurement alone: issues related to the reliability and validity of specific measures, and issues related to the requirements of different stakeholders as will be discussed below. But a central issue is the assumption that services should bring about positive change. Schmidt et al (2000) in discussing the content of outcome measures identify that the treatment of mental ill-health does not necessarily result in an outcome free of symptoms, and in some cases the goal of an intervention may be to maintain current functioning and ensure continuing contact with services.

Moreover, the measurement of change is generally subject to large errors. If an outcome indicator is subject to say 3 per cent error, and the expected range of change in the measure over time is small, the margin of error around the change is very large in relation to the size of the change, and it may not be possible to be certain that change has taken place at all. In the context of education, where the measurement of value added has gone further than in many other fields, Goldstein (1997) argues that value added estimates generally contain too much uncertainty to provide reliable rankings, and therefore at best value added estimates are useful as screening devices to establish outliers, and should not be used as definitive statements of the effectiveness of schools in themselves.

Learning from the review of how to use measures for driving up performance

- a) In order to mitigate against the possibility that any implementation strategy is open to perverse incentives, particular attention needs to be paid to how data are collected (sampling bias), entered (data accuracy), analysed (recognising the need for realistic complexity) and interpreted (the validity of inferences).
- b) It is vital to analyse and try to make sense of data from any measure or assessment process in the light of appropriate contextualising data (much of which is already available) and multiple sources of information.
- c) It is crucial to build systems that allow for integration of data with other sources and for triangulation of results and findings. It is dangerous to rely on any single measure as a stand alone assessment as this can all too easily lead to perverse effects and make it difficult to interpret data.

Section 7: Consideration of implementation issues in England

Method

The following two strands were considered in relation to implementation issues in England:

- a) Consideration was given to the experience of outcome evaluation by CAMHS providers currently. In particular, consideration was given to the experience of members of the CAMHS Outcomes Research Consortium (CORC). CORC represents the main learning collaboration in the UK of child mental health service providers specifically focusing on finding the best ways of evaluating outcomes in child mental health services (www.corc.uk.net). Over half of all service providers in England are members. Consultation was carried out with members of CORC and the central team. It should be noted that the lead author on this report is also Chair and Director of CORC and care was taken to ensure the consultation fed into the outcomes review but did not dictate it. Consideration was also given to the experience of members of the Quality Network for Inpatient Care (QNIC) who have recently started to pilot routine outcome measurement in a small number of inpatient units
- b) Economic analysis of costs of implementation was also undertaken.

Findings

a) Consultation with CORC members

A consultation questionnaire was sent out to CORC members. The questionnaire asked about the strengths and limitations of the measures currently being used by CORC members (Strengths and Difficulties Questionnaire (SDQ), Child Global Assessment Scales (C-GAS), Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA), Commission for Health Improvement – Experience of Service Questionnaire (CHI-ESQ) Goals Based Measure and as voluntary additions for use with children with learning difficulties Child Behaviour Rating Form (CBRF) and Sheffield Learning Difficulties Outcome Measure (SLDOM)). They were additionally asked and if they had used the data gathered in any service reviews.

These responses were then commented on by the CORC central team based on their experience of working with practitioners to implement outcome evaluation.

Consideration of the experience of one CORC membership grouping that had used SDQ value added scores as part of a local PSA agreement (ongoing) was particularly attended to.

- 10 CORC members returned questionnaires

Issues raised are summarised in Table K below, list of comments can be found in Appendix 4.

Table K showing number of responses for each questionnaire

Measure	No of responses relating to measure
Commission for Health Improvement – Experience of Service Questionnaire (CHI-ESQ)	8
Child Global Assessment Scales (C-GAS)	9
Goals Based Outcome Measure	2
Strengths and Difficulties Questionnaire (SDQ)	8
Health of the Nation Outcomes Scales for Children and Adolescents (HoNOSCA)	4
Child Behaviour Rating Form (CBRF)	1
Sheffield Learning Difficulties Outcome Measure (SLDOM)	0
Total No responses	10

Commission for Health Improvement – Experience of Service Questionnaire (CHI-ESQ)

Positives of the CHI-ESQ were reported by 4 members as quick to complete and easy to use and understand. Two members liked the fact that the CHI-ESQ gives space for free text. One member grouping collecting second lots of data found that it can provide useful information on processes which have developed from concerns of service users. Two members liked the fact that it is available for children as well as young people and parents, whilst a few responses stated they liked the fact that it reported on service users’ experiences rather than symptoms.

Some limitations of the measure were reported by 2 members as encouraging positive feedback and giving people little room to say how things could have been improved, particularly on initial questions. One member reported that they find it unusable by non CAMHS services, especially home visits. One member commented that it was difficult that the CHI-ESQ is not available in other languages, while another reported that the CHI-ESQ does not recognise cultural or ethnic differences. One member reported that they find the completion labour intensive.

Comment from CORC central team

It is a popular measure with members, perhaps as it also most often produced positive results. However, some members opt to administer this measure anonymously which tends to elicit less positive results.

SDQ

The SDQ was reported by 4 members to be quick to complete. One member commented positively that it is easy to download and comes in different languages. One member reported that some practitioners find the reports generated from sub-sections useful for diagnostic purposes. Because the SDQs are completed by the service users and not the service themselves, it is thought by one member that the scoring is less likely to be manipulated to show the service in a more positive light. One member reporting finding the SDQ helpful as it looks at three different viewpoints. One member

reported the SDQ to be family friendly, easy to read and good because it can prompt discussions for assessments; it sometimes reminds families to talk about issues they may have forgotten. One member reported that the SDQ also satisfies participation requirements for children and parents.

One member reported that a limitation of the SDQ is that it is confusing to have the different versions for the different age groups. One member commented that families don't like "negatively worded" questions, e.g. family burden. Three members have found the SDQ too broad and do not like the fact that they may have to supplement it with questionnaires for specific disorders, e.g. autism, eating disorders measures. Some members have found the SDQ limited and suitable for behavioural problems, but not necessarily other problems such as learning difficulties and dislike the fact that it does not show representative change scores.

Comment from CORC central team

These measures invokes mixed reactions, although members mostly like it as a general measure and appreciate the different ways in which it can be used; for feedback for the practitioner, using the diagnostic element and the added value score. The issue of "negative wording" in relation to the mention of family burden has been echoed in several implementers' comments voiced at a recent meetings. Members have also expressed concern that it is too broad a measure and may become overused and the effects this may have.

C-GAS

The C-GAS was reported to be quick and easy to complete by eight members. However, two members found it to be a crude measurement of problems which is disliked by many practitioners. One member reporting finding it difficult to achieve inter-rater reliability and so found problems in using it where different clinicians are scoring it at different time points. Two members reported the C-GAS as "arbitrary" as they find it difficult to reflect changes within the specified period. One member reported that it is not very effective in use with children with learning disabilities.

Comment from CORC central team

This measure seems to get quite extreme reactions in that members appear to really like this measure or not like it very much at all, rather preferring the HoNOSCA for use to gain the perspective of the practitioner. Although not usually viewed as suitable for children/young people with learning difficulties, one service recently discussed success with completing the C-GAS as a team for moderate to severe cases, for which they reported a noticeable change in scores between Time 1 and Time 2.

HoNOSCA

One member found the HoNOSCA quick to use. One other member also found that it is good for verifying where there is a need for intervention and identifying areas where there are no difficulties. One member commented positively on the fact that it covers a wide range of problems.

One member commented that a limitation of the HoNOSCA is that it can be confusing and they do not find it very easy to rate only the last 2 weeks, which is the time period specified by the measure. One member reported that the HoNOSCA can be subjective, especially if a different people score time 1 and time 2 due to the re-allocation of cases. One member also commented that you really need to include all questions in an assessment to be able to score it.

Comment from CORC central team

CORC have received far less data on the HoNOSCA than the C-GAS, although the central team are not sure that this is reflective of how much it is actually used. Members who do not like this measure, often cite the reason as being because they feel that is too diagnostically based, however others recommend its use. These differences may reflect the fact that some items within the measures make it is most suitable for adolescents' age groups.

Goals based measures

One member commented that they have found it an excellent for short term solution focused work, and that it fits well with the Choice and Partnership Approach (CAPA) model now being used widely across specialist child and adolescent mental health services (CAMHS).

Comment from CORC central team

The Goals-Based Outcomes measure has been well received by members. The central team do not have very much data on these measures as members are only just starting to use them. The Goals measure seems popular in theory and gets a positive response every time it is mentioned because it is accessible to everyone.

CBRF

One member found this measure useful in gaining a brief overview of behavioural difficulties; they considered the scoring system straightforward and the measure to have short, precise questions. They also commented, however, that the language used in the measure assumed a level of verbal skills and reasoning not appropriate for children with severe LD, and that the language used would not necessarily be familiar to or easy for families/carers to understand either.

Comment from CORC central team

Services have only recently begun piloting this measure, but the comments received by the central team so far received have described it as too lengthy and dense, and felt that one of the sections felt very blaming of parents and their role in child's problem. Another service piloting this measure regarded it as being from a Medical Model of disability and that this does not easily fit with the service philosophy which relates to the Social Model of disability. As such they use it as a useful adjunct to the SLDOM and Goals which they view very positively.

Service Development

Two members responded that they were using data to inform service reviews; whilst another expressed a possibility of doing so once they had submitted their data.

Comment from CORC central team

On the whole, members are not yet at a stage where they are routinely using their data for service development. Against the backdrop of the huge learning curve which is involved in getting outcomes measurement up and running, not to mention the length of time in initiating this process and cultural shift in services, some of the earliest members of the learning collaboration are only now beginning to be in a position where they have data collected and analysed with a view to putting this to use in service development. Members too are understandably tentative about having large enough numbers in their data to be able to make meaningful interpretations, but have begun to look at the findings, triangulate these with other sources of information and taking into consideration the larger context of the service, in order to start to hypothesise about what the data may be actually telling them.

General Comments

Some issues were raised about the fact that whilst the measures themselves are free, the implementation and collation of the data costs services in time and resources. It takes time to collect outcomes data and many services employ research assistants or administrative staffs who are dedicated to the implementation of CORC. Services need someone who is quite influential who can motivate others within the service to collect outcomes data and those services whose commissioners had appropriately resourced this commented positively on this. Services often do not know where to begin collecting outcomes data and require lots of support and advice.

Many services have developed or purchased databases within which they store the scores from the measures. Many NHS services' IT departments will not support certain types of programme or database, which causes a lot of problems for some members.

It should be noted that the National CAMHS Dataset which has been widely consulted on and used by CORC members is now being used by dataset developers across services to determine their databases and allows for the collection of the measures currently being used including the SDQ, HoNOSCA, CHI-ESQ and C-GAS. These measures are now being considered for inclusion in electronic patient records and other data systems being developed as part of national IT developments.

Particular consultation was undertaken with the CORC member grouping involved in local PSA in relation to child mental health based on Parent SDQ scores alone. The member grouping felt positive progress had been achieved and high return rates (33%) had been achieved by introduction of SDQ as a stand alone measure. However, they felt they wanted to supplement this in future with the use of additional measures and other relevant data, as CORC had throughout advised (see consultation response to ODPM via CORC January 2004 - available via CORC website). As a result of their experiences they are proposing to set up a multi-agency CAMHS Information Group with the aim of bringing data together from a range of sources including health,

education, children's social services and youth offending services in one place to provide a rich picture of service activity and outcomes. The Information Group will provide a forum for exploring a range of questions, ideas and hypotheses which will be welcomed as one way of promoting service reflection, development and improvement. This will also allow for an interpretation of the data to be made for the benefit of clinical expert groups, commissioners and service providers and ultimately the children and young people themselves.

Consideration was also given to the experience of QNIC in relation to outcome measurement. QNIC members had chosen to pilot the same measures as those used by members of CORC (C-GAS, HoNOSCA, SDQ) and in terms of service satisfaction were using the CAMHS Satisfaction Scale for Young People in Inpatient CAMHS (CAMHSS). The pilot project had initially included an additional Eating Disorders Examination (EDE) but this was found to be impractical because of its length and has now been optional. QNIC members did not report any problems with the other measures.

b) Consideration of costs of implementation of outcome measurement

There are two sets of costs involved in the collection of outcome data: the cost of collecting the data, and the cost of analysing the data so that it is useful to service providers and to commissioners and funders of services. At present where CAMHS use outcome measurement both parts of the cost are borne by the individual service, and services often employ someone whose job it is to process the data and generate reports.

In the case of both the school-based survey and an instrument for service users we are recommending that a web-based approach should be adopted on cost grounds. This is because the coding and creation of a dataset is automated. There are one-off costs in terms of programming the instrument, but when spread across a large number of respondents these will amount to pence per respondent. In the case of the school-based survey there are administrative costs involved in making arrangements for completion with schools (although these would be small if the survey operated alongside or as an extension to the Tellus survey). Costs should be similar to the costs of Tellus (but considerably less if the two are run together). The processing costs are likely to be small. The main costs will be for analysis and the production of reports, which will vary according to the level of sophistication required, but the production of standard reports by school, local authority or CAMHS area could be automated.

By contrast using paper-based methods data entry costs alone would be around £5 per completed instrument, with any manual coding possibly doubling that. Computer-assisted telephone interviewing would automate the creation of a dataset, but would cost around £10 per head in interviewer costs. Face-to-face completion by a teacher (using supply rates as a proxy) would cost around £7 for a 15 minute instrument. Completion by a generic multi-disciplinary CAMHS team member would cost around £21.50 in staff time for a 15-minute instrument⁵. Thus, for instruments targeted at children and young people it is not cost effective for staff time to be used in taking the child or young person through the questions unless this makes sense in clinical terms,

⁵ Source: Personal Social Services Research Unit (2007) *Unit Costs of Health and Social Care 2007*

in which case the cost would not necessarily be additional. Many CAMHS practitioners already use instruments with children and young people in order to assess the severity of their symptoms. If an instrument were to be used for outcome measurement purposes it might be used more frequently, and rather than merely being part of a young person's record, it would go onto a service database, but there need not necessarily be an additional burden on practitioner time. In the case of CAMHS, where instruments are not routinely used there will be additional costs. However, services which are currently using instruments for outcome measurement report that having the information from instruments helps in service planning and configuration.

There would be some IT costs of linking the completed online instrument both to the individual patient record and to the database where the service-wide information is processed.

The cost of analysing the data and of producing reports that are useful to schools, to service providers and to commissioners does not vary much with increased numbers of respondents. The current situation where each CAMHS service which is using outcome measurement undertakes its own analysis and reporting is a relatively expensive approach. In the state of Ohio, where quarterly measurement of outcomes for both children and adults is obligatory for service providers, all processing and production of reports is undertaken centrally. Each provider receives its own data, aggregated data from other providers in the same part of the state, and aggregated data from across the whole state. This offers opportunities for economies of scale in data handling. As a consequence, the data handling and processing non-reimbursed costs to services are around 1.4 per cent of provider costs (Ohio Department of Mental Health 2008b). Staff time in face-to-face completion of monitoring instruments reimbursed as it is treated as part of the clinical treatment package. On this basis it would be reasonable to estimate that the total cost to services of data collection and handling would be in the region of 3 per cent of budgets (and closer to 5 per cent if each service did its own data processing and production of reports).

Learning from our review of implementation issues:

- a) Online may be the best and most cost-effective mode of implementation wherever possible.
- b) National dataset development IT developments in England, including the development of electronic care records, means that some currently used measures are now being built into future national systems as well as already being built into a range of existing local and regional systems.
- c) Commissioners should allow a minimum of 3-5% of budget spend to ensure outcome evaluation can occur in services.

Conclusions and recommendations

Reminder of remit

To carry out a review of measures in the area of children's psychological well-being and children's mental health, in order to advise the Departments of Children Schools and Families (DCSF) and Health (DH) on which are the best measures that could be implemented from April 2009 onwards, (as part of Public Service Agreements and other means) to influence the system to promote better psychological well-being and effective intervention for those children with mental health problems, including, if appropriate, advising on how to take forward the development of a new approach or measure.⁶

Reminder of Method

Four strands of work were undertaken (June - October 08):

1. Systematic search and structured review of existing outcomes measures in child mental health
2. Consultation with stakeholders about their experiences of using measures
3. Review of use of outcome measures for driving up performance
4. Consideration of implementation issues

Reminder of Methodological Limitations

The tight timescale for this review imposed inevitable constraints. The commission from the DCSF/DH was for the research team to make specific recommendations based on their best understanding of the information that they had been able to review and consider within the timescale. Although the search strategy was entirely systematic, we were only able to obtain a subset of all available materials identified by the review and to consult with stakeholders through existing networks. The findings should therefore be viewed as based on a rapid and critical appraisal of the information that was available to us at the time.

Conclusions

Learning from the systematic search and literature based review of measures:

- a) It was possible to identify a small number of psychometrically robust measures that could potentially be considered for use alongside other measures and data collection, to influence the system to promote better psychological well-being and effective intervention for those children with mental health problems.
- b) However, whilst the number of measures that are available to assess child mental health outcomes is growing fast (and there are some interesting new measures in development), all existing measures identified had limitations as well as strengths and none met the full range of psychometric criteria likely to be essential for successful ensuring they can reliably measure both severity and change over time in key groups. In particular; no measure has been

⁶ In August 2008 the brief was extended to include a consideration of measures of social and emotional skills of children and young people. This aspect of the report is to be completed by December 2008.

tested for bias or differential performance in different UK ethnic minorities and other hard to reach groups and we found no evidence of any single measure being successfully used for performance management of the kind foreseen by this initiative.

- c) It can be difficult to adapt any measure for which the copyright is held by one academic, institution or psychometric corporation to meet a local or particular need, and this potentially puts the academic or institution in a position to “hold hostage” anyone who wants to use and adapt such a measure.
- d) Looking at all existing measures there is a considerable degree of conceptual overlap in terms of item wordings (there may be only so many ways you can ask if someone is feeling sad).
- e) Given the latest psychometric techniques and the level of knowledge now available one possible way forward may be to construct an item bank that can be used to assess different aspects of mental health in children more flexibly in the future without them all necessarily being given all the same items.

Learning from stakeholder consultation:

- a) There is evidence that some of the key measures are now in widespread use and becoming embedded in or alongside systems that can monitor and feedback on outcomes achieved and practitioners report on them positively though recognising their limitations.
- b) It is important to let stakeholders have a degree of choice to use additional complementary measures for their specific populations or contexts.
- c) It is important in the long term to develop systems that allow for flexibility of assessment portfolios and possibility for change over time.

Learning from the review of how to use measures for service improvement:

- a) In order to mitigate against the possibility that any implementation strategy is open to perverse incentives, particular attention needs to be paid to how data are collected (sampling bias), entered (data accuracy), analysed (recognising the need for realistic complexity) and interpreted (the validity of inferences).
- b) It is vital to analyse and try to make sense of data from any measure or assessment process in the light of appropriate contextualising data (much of which is already available) and multiple sources of information.
- c) It is crucial to build systems that allow for integration of data with other sources and for triangulation of results and findings. It is dangerous to rely on any single measure as a stand alone assessment as this can all too easily lead to perverse effects and make it difficult to interpret data.

Learning from our review of implementation issues:

- a) Online may be the best and most cost-effective mode of implementation wherever possible.
- b) National dataset development IT developments in England, including the development of electronic care records, means that some currently used measures are now being built into future national systems as well as already being built into a range of existing local and regional systems.
- c) Commissioners should allow a minimum of 3-5% of budget spend to ensure outcome evaluation can occur in services.

Key recommendations

1) Population survey

Why: To gauge population level of well being/quality of life and access to support.

What: A psychological well being related measure (recommended measure: Kidscreen) to be used alongside or as part of Tellus3 survey (and/or other indicators put in place at the school level)⁷ with the addition of some questions about current access to sources of help (to be drafted), plus the possible use of a measure of emotional and social skills (to be confirmed when the second part of the review is completed in December 08).

How: Such a universal survey could most easily be conducted through computer-based online assessment IT platforms already available in most schools, complemented by a (predominately) web – or mobile phone - administered parent survey.

When: Piloting of this approach to begin in some schools in Academic Year 2009-10 with view to national roll out 2010-11 onwards.

Who: Completed by children over the age of 8 and their parents.

Implementation issues: It is recommended this approach is piloted using a series of increasingly large surveys of representative samples of children and parents to build up knowledge of the technical and data-analysis issues associated with this form of online questionnaire in school. Clear guidelines would need to be established in terms of the potential use of such data by a range of organisations including Ofsted. The implication of possible demands for access to the results using Freedom of Information Act would need to be carefully considered prior to national roll out.

2) Service level outcomes evaluation

Why: To help assess the effectiveness of services and drive up performance.

What: Recommend as a minimum use of parent and child SDQ combined with as two or more additional measures to be determined by services such as goals based measures, measures specific to the presenting difficulties, measures of practitioner views of child and family functioning (such as the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) or Child Global Assessment Scales (CGAS) and measures of outcome of consultation to others as relevant. These may be considered alongside other relevant data including measures of experience of services (such as Commission for Health Improvements, Experience of Service Questionnaire (CHI-ESQ)).

⁷ The review team learnt on the 28th October of the CDCFS/Ofsted consultation on “Indicators of a school’s contribution to well being” which is inviting comments for collation in January 09. Such a consultation is welcomed and the review team suggests the results of this review are shared with the relevant group in DCSF/Ofsted undertaking this important consultation.

How: Mandated via PSA and collected via national dataset collections.

When: From April 2009

Who: Children over 11 and parents of children over 3 accessing specialist child and adolescent services, their practitioners and those consulting such services as relevant.

Implementation issues: Given the complexities of the range of services available it is suggested the measures be mandated in the first instance for use in specialist mental health provision for young people from 3-17 but with a strong recommendation that all services look to use outcome measures as appropriate **(and as a minimum use goal based measures which should be applicable in every service context and with all age groups)**. Commissioners should allocate 3-5% of budgets for implementation of outcome evaluation across services.

3) Future development of a new measure/approach

Why: To develop a new assessment tool consisting of a flexible set of items or scales, that provides greater coverage of the range of key elements and affords a wider measurement range (i.e. able to capture both levels of mental health difficulties and levels of psychological well being and including social and emotional skills); where intellectual ownership is with DCSF/DH (to allow flexibility of adaptation as relevant) and which can be tested with hard to reach groups and possibly used at multiple time points to be sensitive to change.

What: To develop a new measure and linked implementation approach that might include elements of mental health, emotional well being and emotional and social skills and might focus on continuous evaluation (i.e. across meetings rather than at only two time points) with copyright vested in commissioner of measure i.e. DCSF/DH.

When: It is estimated that 2 years will be needed for development and piloting.

How: Via tender with proposed key elements for tender for such a development. The following elements should be included in specification:

- The development of a measure and piloting with hard to reach groups, this should include examination of possible uses for benchmarking using a risk adjustment and added value approach
- To be completed in 24 months
- To be developed by those with expertise in psychometric and risk-adjustment/outcome monitoring research to answer to an external advisory group with relevant psychometric expertise
- Copyright of any new instrument developed to rest with DCSF/DH and IP for the risk adjustment method also
- Item analysis to be carried out to look to include measure of change over repeated measurement.

Who: To be developed by the best group who tender for this.

Implementation issues: DCSF/DH will need relevant academic advisors including psychometric experts and experts in risk adjustment/added-value technologies (IT and statistics experts) to advise them on this process and to review progress of the group who are developing the measure itself.

4) Establishment of a child wellbeing observatory/group

Why: To develop a means to make use of nation-wide data collection on psychological wellbeing and mental health problems to support valid inferences from the data through appropriate analysis and future research and development.

What: This may grow out of existing structures or be separate (DCSF/DH to consider), e.g. might link in with NICE developments.

When: This is likely to require negotiation with current structures.

Implementation issues: The DCSF/DH should work with relevant organisations and emerging groups to help develop meaningful analysis of any data derived, in order to determine whether it is possible to set benchmarks using an appropriate combination of existing or new measures in a contextualised way from 2011 onwards.

5) Communicating with the field

Why: It is important to communicate complexity to avoid being shrugged off as “top down” or “simple minded” and also to share information on the range of available measures to allow all to find the best measures for them.

What: This report and the list of key measures identified should be made available. Generally in communicating to the field (including commissioners, practitioners and service users) it should be acknowledged that this is a complex area and it should be stressed that the DCSF and DH are alive to the complexities and limitations of outcome measurement.

When: From November 2008 onwards.

Who: DCSF/DH and review team members.

How: Via networks e.g. FOCUS, National CAMHS Support Service (NCSS) etc.

Implementation issues: It may be valuable to hold a conference on this in 2009-10 in order to explore further.

Section 9: Summary of proposed way forward for population survey and service evaluation outcomes 2009-10

Goals and objectives	What	Contextualising data	What it can be used for	Future developments
<p>Assess levels of well being and access to services in the community to inform national and local policy and service development</p>	<p>Piloting of survey of children (8 year and over) in school and their parents via online survey: Kidscreen + additional questions re service use + ? measure of social and emotional skills (to be considered once results of review ready in December 08)</p>	<p>DCSF data on pupils including demographics, attainment etc</p>	<p>Along with contextualising data to look at psychological wellbeing across areas and indicate key issues</p>	<p>To develop a DCSF/DH owned measure that can be used to measure aspects psychological wellbeing/ mental health and social and emotional skills in future years</p> <p>Tender to be put out</p>
<p>Assess effectiveness of services and drive up performance</p>	<p>Mandated service level indicator used routinely in all services at intake and following specified period of time.</p> <p>Could be online or paper</p> <p>At first contact and then 6 months later as a minimum (but could be used at more time points)</p> <p>SDQ plus two or more of following domains (as felt relevant by the service)</p> <ul style="list-style-type: none"> - goals based measure - measure specific to the presenting difficulties - measure of practitioner assessment (eg C-GAS or HoNOSCA), - measures of impact of consultation <p>Plus a measure of experience of services (eg CHI-ESQ)</p>	<p>Minimum dataset via NHS</p> <p>National pupil dataset</p>	<p>Benchmarking</p>	<p>Develop approach to triangulate data to allow meaningful benchmarks to be set.</p>

Section 10: References

Atzaba-Poria, N., Pike, A. & Barrett, M. (2004) Internalising and externalising problems in middle childhood: a study of Indian and English children living in Britain. *International Journal of Behavioural Development*, 28, pp.449-460.

Bell-Dolan, D. J., Foster, S. L., & Sikora, D. M. (1989) Effects of sociometric testing on children's behaviour and loneliness at school. *Developmental Psychology*, 25, pp. 306-311.

Bird, S.M., Cox, D., Farewell, V.T., Goldstein, H., Holt, T. and Smith, P.C. (2005) Performance Indicators: good, bad and ugly. *Journal of the Royal Statistical Society, A* 168 part 1, pp. 1-27

Card, N. A., Hodges, E. V. E., Little, T. D., & Hawley, P. H. (2005) Gender effects in peer nominations for aggression and social status. *International Journal of Behavioural Development*, 29(2), pp.146-155.

Conrah D, Sonuga-Barke E, Stevenson J & Thompson M (2003) The impact of maternal mental health and child's behavioural difficulties on attributions about child behaviours. *British Journal of Clinical Psychology*, 42, 69-79

Doucette, A. (2002) Outcomes accountability within child service systems. *Data Matters*, 5, pp. 9 (Georgetown University)

Dougherty Management Associates (2005) *Children's Mental Health Benchmarking Project: Fourth Year Report*
<http://www.dmahealth.com/Reports/Casey%20Yr%204%20Final%20REport.pdf>

Eid, M., & Diener, E. (2004) Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research*, 65 (3), pp. 245-278.

Eid, M., Schneider, C., & Schwenkmezger, P. (1999) Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality*, 13, pp. 283-307.

Fonagy, P., Matthews, R. & Pilling, S. (2004) The Mental Health Outcomes Measurement Initiative: Report from the Chair of the Outcomes Reference Group. London: National Collaborating Centre for Mental Health.

Gardiner, J. (1994) Differential teacher response to type of behavioural disturbance: The good, the bad and the forgotten. (<http://www.aare.edu.au/94pap/gardj94372.txt>, accessed August 2008)

Garcia Armesto, S., Medeiros, H., and Wei, L. (2008) Information availability for measuring and comparing quality of mental health care across OECD countries. *OECD Health Technical Papers No.20*. Paris: OECD

Goldstein, H. (1997) Methods in school effectiveness research. *School effectiveness and school improvement*, 8, pp. 369-95.

Goodhart, C.A.E. (1975) Monetary Relationships: A View from Threadneedle Street in Papers in Monetary Economics, Volume I. Canberra: Reserve Bank of Australia

Holloway, F. (2002) Outcome measurement in mental health – welcome to the revolution. *British Journal of Psychiatry*, 181, pp. 1-2.

Hunter, J., Higginson, I., and Garralda, E. (1996) Systematic literature review: outcome measurement for child and adolescent mental health services. *Journal of Public Health Medicine*, 18(2), pp. 197-206.

Lezzoni, L. I. (2003) Risk adjustment for Measuring Health Care Outcomes. Chicago: Health Administration Press

Lambert, M., & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (ed.). *Bergin and Garfield's Handbook of Psychotherapy and Behaviour Change* (5th Edition). New York: Wiley, 139-193.

Littell, J. H., Popa, M., & Forsythe, B. (2005) Multisystemic Therapy for social, emotional, and behavioural problems in youth aged 10-17. *Cochrane Database of Systematic Reviews* (4), Art. No.: CD004797.
DOI: 004710.001002/14651858.CD14004797.pub14651854.

Marshall, M., Shekelle, O., Brook, R., and Leatheman, S. (2000) *Dying to Know: Public Release of Information about Quality of Health Care*. London: Nuffield Trust

Mayeau, L., Underwood, M.K., & Risser, S.D. (2007) Perspectives on the ethics of sociometric research with children. *Merrill Palmer Quarterly*, 53(1), pp. 53-78.

Merrell, C., & Tymms, P. (2007) Identifying reading problems with computer adaptive assessments. *Journal of Computer Assisted Learning*, 23, pp. 27-35.

Miller, S.D, Duncan, B.L & Hubble, (2004) Beyond Integratoin: The Triumph of Outcome over process in clinical practice *Psychotherapy in Australia* vol 10 , February 2004

Minami, T. and Brown, J. (2008) ACORN Criteria or Effectiveness (ACE) Treatment Outcomes Evaluation Method. ACE Working Group Paper

Neely, A.D. (1998) *Measuring Business Performance: Why, What and How*. London: Economist Books

Ohio Department of Mental Health (2008a) *Ohio Mental Health Consumer Outcomes System: Report 19*.

<http://www.mh.state.oh.us/oper/outcomes/reports/rpt.quarterly.19.pdf>

Ohio Department of Mental Health (2008b) *Outcomes Overview*

<http://www.mh.state.oh.us/oper/outcomes/quick.picks/outcomes.overview.pdf>

Personal Social Services Research Unit (2007) *Unit Costs of Health and Social Care 2007*. PSSRU University of Ken

Pidd, M. (2005) Perversity in public service performance measurement. *International Journal of Productivity and Performance Management*, 54 (5-6): pp. 482-493.

Propper, C. and Wilson, D. (2003) The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy*, 19(2): pp. 250-267.

Raby, C (2007) „Children and Young People’s Vocies in Mental Health and Emotional Well-being. In CAMHS National Cofernce Handbook; Sustaining Comprehensive CAMH Delivery. Available from www.csip.org.uk/silo/files/sustaining-com-camhs-handbook.pdf Last accessed 30th October 2008

Roy, B.V., Veenstra, M., and Clench-Aas, J. (2008) Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre- and late Adolescence *Child Psychology and Psychiatry*

Schmidt, L.J., Garratt, A.M., and Fitzpatrick, R. (2000) *Instruments for Mental Health: a Review Report from the Patient-reported Heath Instruments Group* (formerly the Patient-assess Health Outcomes Programme) to the Department of Health, September 2000.

Smith, P.C. (2005) Performance Measurement in Health Care: History, Challenges and Prospects. *Public Money and Management*, 25(4), pp. 213-220.

Smith, P. (2002) *Measuring Up: Improving Health Systems Performance in OECD Countries*. Paris: OECD

Smith, P.C., Mossialos, E., and Papanicolas, I. (2008) *Performance measurement for health system improvement: experiences, challenges and prospects*. Copenhagen: World Health Organization Europe

Stanger, C., and Lewis, M. (1993) Agreement among parents teachers and children on internalizing and externalizing behaviour problems. *Journal of Clinical Child Psychology*, 22.

Waddell, C., McEwan, K., Shepherd, C.A., Offord, D.R. Hua, J.M. (2005). A Public Health Strategy to Improve the Mental Health of Canadian Children. *Canadian Journal of Psychiatry*, 50 (4).

Verhulst F & Van der Ende J (2008) Using rating scales in a clinical context Chapter in M Rutter, D Bishop, D Pine, S Scott, J Stevenson, E Taylor & A Thapar. *Rutter’s Child and Adolescent Psychiatry*. Oxford: Blackwell, pp 289-298

Appendix 1: Lists of measures Identified by initial database search and consultation with stakeholders (113)

Achenbach System of Empirically Based Assessment (ASEBA)	Missouri Children's Behaviour Checklist
Acting Out Score	Montgomery-Asberg Depression Rating scale for signs and symptoms of depression
Adjustment Scales for Children and Adolescents (ASCA)	Mood and Feelings Questionnaire
Adolescent Adjustment Scale (AAS)	Mood and Feelings Questionnaire
Adolescent TEIQue-SF	My Class Inventory
Adquest	National Checklist for Indicating Psychosocial Problems in Five Year Olds
Assessment of Dual Diagnosis	Ontario Child Health Scale
Beck Depression Inventory	Parent Rating Scales for Children
Behaviour Assessment System for Children	Parent-Child Rating Scale
Behaviour Problem Checklist	Parents Questionnaire
Behavioural and Emotional Rating Scale	Pathology summary score
Behavioural and Emotional Strengths Scale	Patient Health Questionnaire
Belonging - confined to school context	Pediatric Symptom Checklist
Berkeley Puppet Interview Symptomatology Scales	PedsQL Present Functioning
Brief Multidimensional Student Life Satisfaction Scale	Personality Assessment Questionnaire
Center for Epidemiological Studies Depression Scale	Pictured Child's Quality of Life Self Questionnaire
Child Adaptive Behaviour Inventory	Piers-Harris 2
Child and Adolescent Functional Assessment Scale	Positive and Negative Affect Scale
Child Behaviour Rating Form	Practical Adolescent Diagnostic Interview
Child Depression Inventory	Prosocial Behaviour Questionnaire
Child Health Questionnaire	Psychiatric Assessments Schedules for Adults with Developmental Disabilities
Child outcome rating scale	Psychological General Well-Being Index
Child Problem Behaviour Screening Test	Pupil Evaluation Inventory
Children's Global Assessment Scale	Quality of Life Enjoyment and Satisfaction Questionnaire
Children's Moods, Fears, and Worries Questionnaire	Quality of Life Profile Adolescent Version
Children's Pathology Index	Reiss Screen
Children's Social Behaviour Questionnaire	Reynolds Adolescent Adjustment Screening Inventory
Classroom Behaviour Inventory	Rochester Adaptive Behaviour Inventory
Conners Rating Scales	Rosenberg Self-Esteem Measure
Culture-Free Self-Esteem Inventory (2nd Edition)	Rutter Scales
Development and Well-being Assessment	Schedule for Affective Disorders and Schizophrenia for School-Age Children

Developmental Behaviour Checklist	Schedule for the Evaluation of the Individual Quality of Life: SEIQoL
Devereux Early Childhood Assessment	School Children's Happiness Inventory
Devereux Elementary School Behaviour Rating Scale	Self Report Questionnaire
Devereux Scales of Mental Disorders	Self-Perception Profile for Children
Diagnostic Assessment for the Severely Handicapped-II	Social Competence and Behaviour Evaluation Scale
Diagnostic Interview for Children and Adolescents	Social Competence and Behaviour Evaluation Scale
Dominic	Social Inclusion Survey - peer rated
Early Childhood Inventory-4	Solution Focused Recovery Scale
EQ-5D - health-based	Story stem narratives
Eyberg Child Behaviour Inventory	Strengths and Difficulties Questionnaire
Feelings, Attitudes and Behaviours scale	Strengths and Weaknesses in ADHD and Normal Behaviours
Five to fifteen	Symptom Check List
General Health Questionnaire	Symptoms and Functioning Severity Scale Peabody Treatment Progress Battery
Generic Children's Quality of Life Measure	Target Symptom Rating
Global Assessment of Functioning Scale	Taxonomy of Problem Situations
Global Assessment of Psychosocial Disability	Teacher-Child Rating Scale
Guess Who - peer rated	Tell Me A Story
Hamilton Rating Scale for Depression	The Child Symptom Inventory
Health and Behaviour Questionnaire	TNO AZL Children's Quality of Life
Health of the Nation Outcome Scales	Washington Symptom Checklist
Index of Child and Youth Well-Being	Werry-Weiss-Peters Activity Rating Scale
Internalizing Symptoms Scale for Children	World Health Organization Quality of Life Assessment
Kidscreen	Youth Outcome Questionnaire
Life in School Checklist	Youth Quality of Life Instrument- Research Version
Massachusetts Youth Screening Instrument- 2	Young Person's Child Outcomes Rating Evaluation
Me and My School	

Appendix 2: Summary of quality assurance processes for systematic search

Table showing quality assurance processes in place stages 1-5

Aspect of review process	Methods used	Groups consulted	Comments
1. Creating search terms	<ul style="list-style-type: none"> ○ Discussion to reach consensus 	<ul style="list-style-type: none"> ○ Individual ○ Review group ○ Overseeing group 	Search terms generated by review group, approved by overseeing group
2. Screening measures	<ul style="list-style-type: none"> ○ Traffic lights <ul style="list-style-type: none"> ○ Red ○ Amber ○ Green ○ Discussion to reach consensus 	<ul style="list-style-type: none"> ○ Individual ○ Review group ○ Overseeing group 	Traffic lights were used such that measures that clearly fitted the search criteria received a green light, those that satisfied exclusion criteria received a red light, those that were ambiguous received an amber light and were discussed first with the review group and second with the research group (if first round of discussion did not result in resolution)
3. Deciding psychometric criteria	<ul style="list-style-type: none"> ○ Discussion to reach consensus 	<ul style="list-style-type: none"> ○ Review group ○ Overseeing group ○ Expert groups within overseeing group - Psychometricians 	Criteria outlined by expert group, operationalised by review group and commented by overseeing group
4. Deciding implementability criteria	<ul style="list-style-type: none"> ○ Discussion to reach consensus 	<ul style="list-style-type: none"> ○ Review group ○ Overseeing group ○ Expert groups within overseeing group - Clinicians/education experts 	Criteria outlined by expert group, operationalised by review group and commented by overseeing group

Appendix 3: Stakeholder consultation: summary of comments

Qualitative feedback phase 1

BYI

Strengths

1. Usability: Easy to use; used in routine clinical practice; quick
2. Current ways of using: used in routine clinical practice; good tool for pre and post measure; normed; excellent screening tool; good to help with formulation; good for baseline monitoring; Helpful for use when assessing depression; encourages discussion; appropriate for older CAMHS
3. Description of Scale: Good age range; encompasses five scales in one measure (multifactorial); Looks at range of emotional difficulties rather than just one

Weaknesses

1. Financial cost: Expensive tool
2. Criticisms of Scale: Restricted age range; Open to bias; May be too focused on clinically significant problems rather than level of functioning
3. Reporters: -just looks at child, nothing objective from parents and teachers
4. Usability: Discussions indicate that it would be useful for all CAMHS clinicians to use with appropriate cases; not to be used in isolation

ASEBA [CBCL]

Strengths

1. Usability: Easy to use
2. Current ways of using: Used as part of initial assessment because covers a broad range; used in research project not sure how useful clinically
3. Description of Scale: Comprehensive: both internalising and externalising measures; can compare child, parent, school

Weaknesses

1. Financial cost: Not used routinely -costly
2. Usability: Difficult and time consuming to score
3. Criticisms of Scale: Lengthy; not used as before and after; Found it to be reductionist of rich data without benefit

CORS

Strengths

1. Usability: User friendly; Straightforward
2. Current ways of using: It is most feasible for child and practitioners and promotes reflection for child and practitioner; Integrated into therapy process and values the 'consumers' perspective-gives you as clinician direct feedback about your performance, so promotes therapeutic dialogue; I have done a qualitative research using focus group discussions with children and young people. The measures have affected the way I practice in CAMHS in many levels.
3. Description of Scale: Really good self report measure from the child's perspective; Measures global distress and is validated; Useful before and after

Weaknesses

1. Criticisms of Scale: Brief-not diagnostic and not focussed on symptoms; not objective

Conners

Strengths

1. Usability: Easy to use and score
2. Current ways of using: Good for ADHD and developmental disorders; Long version is more useful to use; Ok for baseline information; Used in context of family systems;
3. Description of Scale: Tools available for teacher parent and child; useful oversight from parents and teachers and also an option for child; ADHD focus; Long version for assessment and short for monitoring;

Weaknesses

1. Financial cost: Expensive to buy
2. Usability: Long version-time consuming; long and complicated-needs training; Less useful tool in assessment
3. Criticisms of Scale: Too specific; ADHD specific; Does not look at other aspects (besides parents, teachers and child); Tend to over report difficulties as parent's positively respond to items; Overly negative; Bit of a tautology; Don't like overlap of scales; Would be far better if ADHD not mentioned -too narrow-guides people inappropriately-only screening but used as a diagnostic tool

HoNOSCA

Strengths

1. Usability: Easy and quick to use
2. Current ways of using: Good tool to measure any changes; Depends on clinician does not need to be sent out to anyone; clinician rated; Can act as a prompt when assessing; Useful in combination with other scales; Good for comparison across cases and across primary and secondary care.
3. Description of Scale: Some useful questions; Not relevant for kids with learning disabilities; The pre and post scores can measure outcome; Diagnosis led

Weaknesses

1. Usability: Needs training; not across settings e.g. schools Hard to use as outcome - don't necessarily ask all these questions. Too psychiatric, not useful for younger children; time consuming; Most MDT members dislike this measure- Reluctance of some professional groups to complete; Rather long and detailed; use tailed off in clinic especially following due to limited feedback
2. Criticisms of Scale: Subjective-clinician rated; Not learning disability friendly; Some areas over encompassing; Too broad; Seems to minimise complexity of problems; Does not give meaningful data; Not really meaningful -could not capture change in complex populations e.g. LAC; Less about emotional well being of the child; High scores can indicate unknown information-not found popular with clinicians so no longer used; Not always appropriate;

MACI

Strengths

1. Current ways of using: Good for complex cases, Good for adolescents, Only for specific cases -not routine
2. Description of Scale: Covers broad spectrum of difficulties, Very detailed, gives an in-depth perspective offering a good overview of the emotional well-being of the young person, often clients respond that the impression is spot on in terms of describing them.

Weaknesses

1. Financial cost:
2. Usability: Too long to be used as a routine tool; Time-consuming, too in-depth and hand scoring is time consuming; Need high motivation from young person; Laborious; only applicable for older adolescents.
3. Criticisms of Scale: Too specific; ADHD specific; Does not look at other aspects (besides parents, teachers and child); Tend to over report difficulties as parent's positively respond to items; Overly negative; Bit of a tautology; Don't like overlap of scales; Would be far better if ADHD not mentioned -too narrow-guides people inappropriately-only screening but used as a diagnostic tool

SDQ

Strengths

1. Usability: Used in routine clinical practice; widely used and well validated; Know its reliable Easy-clinicians can do; easy to score; Found most useful when used more than one version
2. Current ways of using: used as part of triage process when processing referrals; When put in a graph it is really helpful for families to explore needs, strengths etc; Excellent-free data for untreated; Useful supplemental scale for comparison; Good when used with HoNOSCA; Only for tier 3-not tier 4; Used for interventions-mandatory
3. Description of Scale: Good age range; Very good pre and post measure; Before and after; Good outcome measure :asks different people and self for a perspective; child and parent ratings; Helpful to compare different versions; brief; quick; Comprehensive overview of child's difficulties

Weaknesses

1. Usability: Difficult to get follow-ups completed; Don't always get completed post measure forms back from families; Can be mistaken for a diagnostic tool and a 'screening' out tool if used irresponsibly; Used extensively by other services so kids and parents becoming fed up by continually filling in the same form
2. Criticisms of Scale: Ratings can be confusing to interpret i.e. pro-social scale; May not be sufficiently sensitive to change; Tends to focus more on negatives; over reports difficulties especially parent version; scoring can be time consuming; only focus on symptomatology; very poor scale; feels too broad; sometimes too simple; not particularly sensitive for complex client groups; very subjective -used too generally; ratings can be so variable according to user understanding; risk of delivering satisfaction not outcomes if only part is used e.g. only parental SDQ; a bit crude; Problems measuring short term (3 sessions) interventions; Although cross-culturally evaluated we found difficulties across measures

Qualitative feedback phase 2

ASEBA

- We have only just purchased this and haven't used it yet
- Its very long to administer and not good for teenagers. Self harm, eating disorders or trauma
- 360 degree view helps

BYI

- Would tend to Beck depression inventory with bright young people or Berleson
- More useful if using CBT as an intervention. Children do not always know how they feel so get a lot of false negatives
- Useful for target groups but limited breadth for whole service

HoNOSCA

- I was at a pilot site – I found it unwieldy, that it did not fit the children attending this tier 2/3 service very well and lengthy to complete
- Doesn't reflect severity of some specific but debilitating difficulties
- This used in every case but if the case is re-allocated and therefore a different clinician doing post measure there may be inconsistency in subjective scoring. Only valid if you ask all the questions requiring a score which isn't always necessary
- Crude measure but sensitive to changes
- Expensive to use in community; rater reliability/drift. 'Medical'/diagnostic assumptions vs narrative/problem formulation
- Essentially a staff view

SDQ

- Highly flexible and adaptable – better for younger children but youth versions now available.
- Very quick & easy to score – reports can be produced online
- This is ok for behaviour problems but doesn't give a good measure for other things. It's too simplistic. There have been many cases where a parent/child report an improvement but when SDQ's are scored there is no change. SDQ was never developed as an outcome measure.
- Good for screening
- Hard to ascertain if this is an accurate measure – and evidence of national/county norms would be really helpful/Spreadsheet reporting system to/from CORC really hopeless but the SDQ basic measure could have fantastic potential – if implementers supported with sensible tools. Practitioners really resistant to using it – because of the time taken (not necessarily a valid concern) so hard evidence of it's effectiveness would be so compelling. The tools provided are inadequate for this.
- Easy to administer and families do not have any difficulty with it however it is limited in being able to measure change. Not found to be very useful with psychiatric conditions, eating disorders, trauma, self harm and teenagers
- This and other measures were very poorly presented in the initial DoH Framework of Assessment manual. Marking systems were sometimes confusing
- Easy to use especially with Computer print out of results.
- Used for research only

- Best of limited bunch. Wide applicability but not universal. ONS Added Value score is v valuable. User/ carer rating is better (than eg HONOSCA). One size cannot fit all economically unless Subjective (eg User target / goal-based). Using for service evaluation
- Difficulty in implementing this measure with current IT system as measure is copyrighted to YIM database, means inputting data duplication into two systems or amalgamating two reports

Appendix 4: Comments from CORC member consultation

CGAS

Strengths

- Clinical impression
- Easy, quick, fair
- Quick and easy for clinicians to complete
- Quick indication about the level of functioning- not costs which is good.
- Used across all tiers and services. Very easy to use Easy access to on-line training.
- Clinician rated. Quick and easy
- Quick measure
- Quick to complete

Weaknesses

- Risks encouraging the delivering of services through threshold scores –i.e. if they score 80+ then need a service, any less and they don't get it.
- Needs training to get consistency and performs poorly for children with learning disability
- Some clinicians feel that it doesn't say much.
- Not appropriate for short term solution focused interventions. Not easy to embed in systems as the measure has restrictive copyright
- Mid bracket rating can be somewhat arbitrary
- Crude "construct"
- Crude measurement of problems. Disliked by clinicians. Difficult to achieve inter-rater reliability so problems in using as an outcome measure where different clinicians scoring. This is often the case especially where CAPA is used.

Other comments

- Trying to use to evidence pre-post intervention improvements.
- Failure to consistently gather data makes it very difficult to make comparisons. Identifying that close contact frequently lost with vulnerable children/families (i.e. can't get follow up data and suspect the reason is that mainstream professionals do not have or prioritise a relationship where they can ask/obtain it
- Could be a great tool. Poorly presented, cumbersome data collection and poor reporting makes this a real struggle to get mainstream support. Identifying that a whole host of other follow up data are not recorded either
- Using in outcome study funded by MRC – services involved using for service review
- We are looking at adapting the CGAS to be part of the threshold criteria to our service and add this to our systems/processes. This will enable us to map the type/tier of referrals coming in across the county and identify hotspots. We will also have an easy referral criteria that for GP's to reference. Future developments will look at how this can be adapted to form part of an electronic referral template into our services.
- Research
- SOFAS may be better – not tied to symptoms

CHI-ESQ

Strengths

- Potentially great tool for measuring other interventions
- Easy to use, available for children as well as young people and parents. Easy to score and easy to audit.
- Provides useful information on process – developed from concerns of service users
- I think it is a very good measure especially for adolescents as it is not about symptoms but about the experience of being listened to and cared for. It is short. Little costs involved, can be used in different settings etc.
- This measure is easy to use, and age appropriate. It enables young people and their parents/carers to add narrative on how they found the service and the service to then pick up any issues and be pro-active in improving service delivery
- Quick to complete. Easy to understand. Gives space for free text.

Weaknesses

- Unusable by non CAMHS service
- Not in other languages.
- Limited in what it measures. Therefore outcome measure doesn't always reflect on what change/improvement has occurred. OK for behaviour problems but not others.
- Subjective, especially if a different person scores post measure due to re-allocation of case. Need to include all questions in assessment to be able to score.
- Only measures satisfaction – important but not necessarily correlated with outcome
- Not capturing symptoms, (which can also be an advantage)
- We have adapted and drafted a version of the CHI-ESQ to use in in-patient T4 service that also incorporate Healthcare Standards. This is currently being piloted
- Orientated towards clinic based consultations – not home visits

Other comments

- Only in potential. Very frustrating
- Labour intensive completion. Simple software tools could make the data collection more effective and easier to analyse
- Yes the qualitative data (open ended questions) was useful to gain an understanding of what service users think.
- Using in outcome study funded by MRC – services involved using for service review
- Yes, we use the 'free text' part of the report on expressions of satisfaction and report to the Trust Board monthly. We use the 'free text' part to identify areas of concern in the service and act upon issue raised, this is reviewed monthly by the management team
- We have added some extra questions that will evidence us meeting the Healthcare Standards for CAMHS. This means we have a constant source of data and negates the service having to run additional audits through out the year

- We have used the evidence in these measures to publish some posters, detailing young people's experience of service and improvements made, the aim is to encourage people to 'have their say' through the CHI ESQ and help shape the future developments of the service
- Data not analysed as yet

HoNOSCA

Strengths

- Free to use and score. Covers a wide range of problems
- Use in the Tier in-patient unit. This is part of a pilot project
- Quick

Weaknesses

- Subjective, especially if a different person scores post measure due to re-allocation of case. Need to include all questions in assessment to be able to score.
- Too early to provide any feedback on this measure
- Does not measure all constructs of interest

Other comments

- Used as part of research

Goal based Outcomes Measure

Strengths

- Used in Primary Mental Health Team, YOS and LAC team

Weaknesses

- None listed

Other comments

- Excellent for short term solution focused work, fits well with the CAPA model

CBRF

Strengths

- We are using this as pilot in our service. Useful in gaining a brief overview of behavioural difficulties, short, precise questions (for a clinician to complete) and straightforward coring system

Weaknesses

- Not easy for families and carers to complete – a bit wordy not easy to follow if not familiar with this type of language. Some of the questions are not applicable to children with severe LD – they assume a level of verbal skills and reasoning. Some of the questions would automatically score 'high' for most children with LD e.g. poor concentration skills.

Other comments

- We do think it is really useful to measure outcomes, however in LD progress is usually slow and change is minimal. It can be quite difficult to 'prove' that our input has helped.

SDQ

Strengths

- Family friendly, easy to read, prompts discussions for assessments, reminds families to talk about issues they may have
- Service user led and so less vulnerable to “gaming” (or at least games that show the service in a good light”).
- Easy to use and score and comes in different languages. Free to download and use.
- Good outcome measure and easy to understand and score with the use of the Youth in mind database.
- Used in Community Tier 3 work
- Versions for parents/teachers and child
- Quick – only 25 items
- Quick to complete. Some clinicians find the reports generated from sub-scores useful for diagnostic purposes.
-

Weaknesses

- Multiple versions complicated for practitioners to keep on top of. We had to retype them because the web site only provides poor quality, unsuitable paper copies that had to be adapted to drop words like “clinic”
- Broad general measure – may need supplementing with specific measures for some disorders (ie eating disorders, autism)
- Limited in what it measures. Therefore outcome measure doesn’t always reflect on what change/improvement has occurred. Ok for behaviour problems but not others.
- Not appropriate for short term solution focused interventions. Not easy to embed in systems as the measure has restrictive copyright
- Not sensitive to change. Classifies all children as hyperactive
- Self completion version only validated for use for over 11’s. Would be useful to have a measure which covered a wider age range or use two measures to obtain views from younger children. Confusing to have a parent version for 3/4year olds, 4-16, 4-17. and Self completion version for 11-16 and 11-17. For parents - Under 5’s measure and then 5-17 would make more sense in aligning measures to the way services are organised. Negative wording may be disliked by some families – e.g. asking how much of a burden the child is on family. Orientated to Identified child so not taking a systemic view of problems.

Other comments

- Trying to evidence pre-post intervention improvements. Failure to consistently gather data makes it very difficult to make comparisons
- Gives an indication of risk
- Identifying that close contact frequently lost with vulnerable children/families (i.e. can’t get follow up data and suspect the reason is that mainstream professionals do not have or prioritise a relationship where they can ask/obtain it
- Identifying that a whole host of other follow up data is not recorded either
- Could be a great tool. Poorly presented, cumbersome data collection and poor reporting makes this a real struggle to get mainstream support

- Using in outcome study funded by MRC – services involved using for service review
- If the SDQ is the PSA 12 core measure adopted, future developments will be planned in Lincolnshire to address the duplication/information sharing across all children's' services with regard to the SDQ.
- There will be a big learning curve to address with all children's services with respect to the timeline of implementing the SDQ, duplication and validation issues around this measure
- Best when combined with Teacher / Parent versions
- Data not analysed as yet.

Other Suggested Measures

- PANSS

High reliability – sensitive to change, Required training, Used in Research
Needs reworking for UK/EIP population

- FACE – Youth Version

Web-based reporting available, Few data at present, Being redeveloped with new risk assessment system built in.

Appendix 5: Potential conflicts of interest

In interests of transparency all those involved in this research were asked to identify any potential conflicts of interest. These are listed below:

Miranda Wolpert is Chair and Director of CAMHS Outcome Research Consortium (paid directorship 1.5 days a week)

Miranda Wolpert, Peter Tymms, Pam Meadows, Peter Fonagy, Panos Vostanis, Norah Frederickson and Neil Humphrey are all part of the Targeted Mental Health in Schools National Evaluation which has developed its own measure for assessing mental health in schools (the Me and My School Measure)

Jeb Brown is an independent consultant and statistician who has been involved in questionnaire development but never received and royalties or payments for any copyrighted measure and comments he does not plan to in the future. He points out that as an independent consultant it is in his interest to be helping to develop new and improved questionnaires rather than relying on existing copyrighted measures.

Michael Rutter has developed questionnaires that are used in Mental Health. In particular the Rutter Scales.

Crispin Day is also involved in developing measures of child outcomes and experience of service including the Child and Adolescent Service Experience questionnaire (ChASE). Crispin Day acts as Chair of the Monitoring, Evaluation and Accountability SubGroup, National CAMHS Review.