



Therapy outcome measures: Where are we now?

Alexandra John

To cite this article: Alexandra John (2011) Therapy outcome measures: Where are we now?, International Journal of Speech-Language Pathology, 13:1, 36-42, DOI: [10.3109/17549507.2010.497562](https://doi.org/10.3109/17549507.2010.497562)

To link to this article: <https://doi.org/10.3109/17549507.2010.497562>



Published online: 15 Feb 2011.



Submit your article to this journal [↗](#)



Article views: 10536



View related articles [↗](#)



Citing articles: 8 View citing articles [↗](#)

Therapy outcome measures: Where are we now?

ALEXANDRA JOHN

University of Sheffield, Sheffield, UK

Abstract

Outcomes information contributes to the provision of quality services: sharing that information requires speech-language pathologists (SLPs) to use terminology readily understood by professions ranging from health and education to social and voluntary services. The Therapy Outcome Measure (TOM) provides a way of presenting outcome data in a digestible form, comprising part of a range of multiple measures used to collect information on the structures, processes, and outcomes of care. TOM was developed to provide a practical method of measuring outcomes in routine clinical practice. Furthermore, it has been used in a number of research studies as an outcome indicator. As an example of its utility in research, the article cites a benchmarking study, together with examples of internal and external benchmarking of outcomes intended to illustrate how the benchmarking of TOM data can inform practice. The TOM can therefore inform SLPs on their own outcomes, the outcomes for specific client groups, and, by benchmarking TOM data, can contribute to the delivery of better, more efficient services.

Keywords: *Speech-language pathology, outcomes, therapy outcome measures, benchmarking, language delay, language disorder.*

Background

Speech-language pathologists (SLPs) aim to provide good quality services to children and adults who use them, providing the right services at the right time and delivering the best effect within available resources. A clear and objective understanding of what constitutes “quality” is necessary in order to evaluate the success of these services and to improve them over time. One definition of quality is that a quality service should be: effective in using treatments of known worth; efficient in making best use of resources; responsive in meeting the needs of the individual and their carers; and equitable in providing fair and equal access according to health need (Department of Health, 1996). The ability to provide a quality service requires SLPs to have access to relevant information which informs their own professional practice and informs other professionals who interact with SLP services, the users of the service, the service managers, and, not least, those who commission and fund the services. By definition, relevant information requirements, including in respect of the structure, process, and outcomes of services, will differ depending on the specifics of each case, and will reflect the different perspectives of stakeholders. Service users may judge quality on whether their own needs have been met and whether

their outcomes relate to satisfaction with the care provided and whether the problem has been attended to, resolved, reduced, or eliminated. SLPs judge quality on the extent to which processes and outcomes of care meet the perceived needs of service users, with good outcomes and goals achieved from the SLPs’ perspective. Managers, for their part, judge quality on whether efficient and effective care is provided to the greatest number of people within available resources, with good outcomes delivered and goals attained at minimum cost. The commissioners who fund the services, in turn, want information on the quality of services provided and the cost-effectiveness of those services. In short, they want to know that the funded services make a difference (National Health Institute for Innovation and Improvement World Class Commissioning, 2009), albeit that making a difference can itself be difficult to measure. From an historical perspective, a long-held medical maxim has been “to do good and to do no harm” (Bauman, 1991, p. 9), with the onus resting on the clinician to determine the individual’s unique needs and to provide the most suitable type of intervention, its intensity, and duration. It is, thus, a central requirement of the provision of a quality service that all recipients of healthcare receive appropriate, effective, and timely care, which meets their expectations. Despite this, the *de facto* standard

of healthcare received by patients, in the UK at least, has varied greatly depending on such factors as the clinician's knowledge and on the availability and extent of local resources, services, and finances; it has not, as a consequence, always fulfilled the needs or expectations of the patient (Swage, 2000). Therefore, the evaluation of quality requires a decision on what to measure, from whose perspective, what form of measurement to use, who should undertake the measurement, how to analyse data, and, finally, how the information gained will be disseminated and acted upon. One means of addressing the various factors contributing to an evaluation of quality is via the use of appropriate comparisons and through an awareness of the nature of the variations observed. For example, there may be comparisons against established objectives, standards and guidelines, or against patient needs and expectations. Comparisons can then be performed within a health service, or can be used to compare results with a similar service or services, evaluating quality in one area against that in another. In order to facilitate an evaluation, quality score cards have been developed for use in the UK National Health Service (Department of Health Primary and Community Team and Primary Care Contracting, 2009; National Health Institute for Innovation and Improvement, 2010). The use of the "balanced scorecard", as shown in Table I (Stevenson & Spencer, 2002, p. 92) is one approach to identifying specific areas of information upon which providers can assess, reflect, and report. Pam Enderby's work has contributed to providing information on each level of the balanced scorecard. She has investigated ways of identifying the effectiveness and efficiency of SLP intervention and, where measures were lacking, she has worked to develop valid and reliable methods of capturing the required information. At the heart of this work has been the drive to improve levels of care and to provide accountability. The contribution provided from within Pam Enderby's body of work represented by the Therapy Outcome Measures (TOM) is the focus for this article.

This article provides examples of the employment of the TOM, summarizes briefly the work of Pam Enderby in developing it, and demonstrates how the TOM, as an outcome measure, has informed practice.

Development of the TOM

Outcomes are defined as results or visible effects (Oxford English Dictionary). In the context of healthcare, outcomes should reflect change resulting from interventions, and represent a key aspect of assessing the effectiveness of care. Frattali (1998) identified six distinct health outcome types: *clinically derived* outcomes (based on a specific clinical condition); *functionally derived* outcomes (based on the functional performance of a given condition);

socially derived outcomes (related to social participation); *client derived* outcomes (predicated on perceived quality of life and/or satisfaction); *administratively derived* outcomes (including waiting times and contacts); and *financially derived* outcomes (the cost effectiveness of care). Measuring these different outcomes requires differing outcome measures. There is no single approach that can be adopted across the spectrum, but rather a number of measures need to be used concurrently and the data interpreted to inform practice.

During the 1980s and 1990s, Enderby worked to develop methods of recording information to inform practice. This work covered a number of areas relating to the development of methods for recording structure, process, and outcome information. Enderby worked on the development of the SLP codes for the Read coding system used in the UK National Health Service. Read codes aimed to record electronic data entry of patient information to facilitate audit and research. Part of this work involved developing a measure of outcomes for SLPs that would capture the changes effected as a result of therapy intervention. At that time, outcome measurement focused on achieving treatment goals or used the results of standardized assessments; these measured levels of impairment or communication. As a consequence of this, other aspects of intervention, including social and emotional issues, were frequently excluded from outcome measurement. Comparison of outcomes was complicated by the fact that different measures were adopted in routine clinical practice, according to the condition, severity, and age of the individual. In essence, the commonly-used, goal-based outcome measures were intended to provide information on the gains or benefits resulting from treatment. However, it is now acknowledged that goals achieved provide only part of the picture, as achieving a goal can result in different outcomes for different people, reflecting the complexities of capturing the outcomes of an intervention and assessing holistically the outcome of care (Enderby, John, & Petheram, 2006). It is also difficult to concatenate an individual's goals data. In reviewing the case notes of over 300 cases, Enderby (1992) found that the goals of therapy could be classified under four descriptions: to identify and reduce the disorder/dysfunction; to improve or maintain the function and ability; to assist in achieving potential or integration; and to alleviate anxiety or frustration. The first three goals related to the three dimensions of the World Health Organization's (WHO) International Classification of Impairment, Disability, and Handicap (ICIDH) (WHO, 1980). The ICIDH classification system was based around the body's structure and function, the individual, and society, which could be individually rated to denote the level of difficulty experienced. A status of *normal* was defined as normal for a human being given their age, sex, and culture, and any

difficulty reflected the deviation from this status for that individual. Any *difficulties* experienced could impact on the individual's well-being. The TOM used the ICIDH conceptual model to provide a multiple measure of outcome. The TOM dimensions initially comprised impairment, disability, and handicap, to which Enderby (1992) added that of well-being/distress, in order to capture emotional status. The names of the TOM dimensions were changed in 2001 to reflect the revised version of the IDIDH (International Classification of Functioning Disability and Health, ICF, WHO, 2001), while the conceptual basis of each dimension has remained unchanged. The definitions for each TOM dimension are shown in Table II (Enderby & John, 1997; Enderby, John, & Petheram, 1998; 2006).

Each of the TOM dimensions is rated individually. The 1992 TOM core scale used a 6-point ordinal rating scale, with 0 representing the severe end of the scale and 5 representing *normal* for a human being given age, sex, and culture (Enderby, 1992). The integers were defined with a semantic operational code that defined the severity of the difficulty experienced on each dimension. Since its inception, the TOM has undergone rigorous development work. Part of the development between 1992–1996 included the introduction of client-specific descriptors and the use of undefined half-points to increase the sensitivity of the measure. A TOM manual for SLP was published in 1997 and in 1998 for physiotherapy, occupational therapy, and rehabilitation nursing; these manuals included client-specific scales. Following a large benchmarking study, a second edition of the TOM was produced which incorporated all the client-specific scales for use by rehabilitation professions, including SLP, physiotherapy, occupational therapy, rehabilitation nursing, and hearing therapy (Enderby et al., 2006). The manual sets out the development of the tool and its introduction and use. It describes: the core scale; 32 client-specific scales, with six defined integers, with

the provision to use undefined half-points that together provide an 11-point ordinal scale; a data collection sheet; guidelines; the coding system; and the computer program for analysing the data (Enderby et al., 2006). The data sheet allows the recording of information on SLPs, an individual's name and date of birth, location of treatment, the number of contacts, the duration of treatment, the aetiology code, the disorder code, TOM admission, and intermediate and final rating at time of discharge. A TOM Data Analysis Program was developed by Brian Petheram and his team between 1993–1994 at the Speech and Language Therapy Unit at Frenchay Hospital in Bristol, UK, and has been used by Enderby and her research team in TOM data analysis. The combining of the TOM manuals in 2006 into one manual aimed to provide a common language for rehabilitation professionals to use together, particularly when working in teams. The rater is able to use their favoured assessments, professional judgement, and reports in order to decide on the TOM rating and to provide a quantitative number to qualitative information at set points in intervention. This allows change to be captured, whether that change is positive, negative, or sustained. By capturing each aspect of intervention, the outcomes should better reflect the outcomes of care. Figure 1 shows the factors considered within each of the four TOM dimensions.

Introducing any outcome measure requires support from senior management, as staff time is needed to learn its use. Staff time is also required to implement the TOM, to develop consistency in rating, and to integrate its use into existing routine data collection and analysis systems. Those staff

Table I. Balanced score card (Stevenson & Spencer for King's Fund, 2002). Reprinted with permission.

Balanced score card	
Dimension	Evaluation
Cost effectiveness	<ul style="list-style-type: none"> Quality of life measures Satisfaction surveys Complaints
User satisfaction	<ul style="list-style-type: none"> Goal achievement Health/social benefit Appropriate discharge
Outcomes	<ul style="list-style-type: none"> Intervention Treatment/therapy Care pathways Timing of care/referral
Process	<ul style="list-style-type: none"> Skill mix Access
Structure	<ul style="list-style-type: none"> Length of treatment Reduction in care package

Table II. Therapy outcome measures four dimensions.

Impairment	Problems in body function or structure such as a significant deviation or loss.
Activity	Difficulties an individual may have in the performance of activities.
Participation	Problems an individual may have in the manner or extent of involvement in life situations.
Well-being	Emotional effect resulting in an upset, distress, or satisfaction with status.

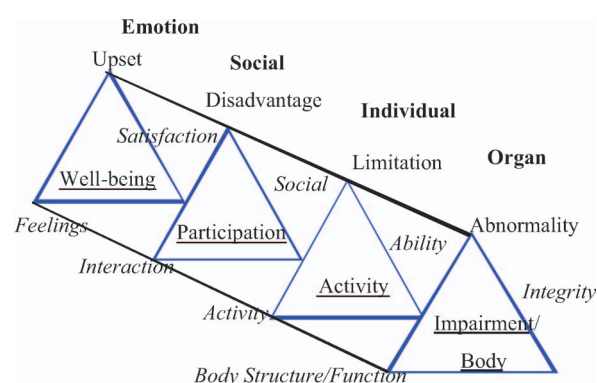


Figure 1. Therapy outcome measure dimensions.

using the TOM need to feel comfortable using it and confident that they are using it reliably, particularly if concatenating and comparing TOM data across services for benchmarking. It has therefore been recommended that TOM training incorporates a training session, followed by individual rating of 10 cases, and then team consensus rating on the TOM for pre-post case rating facilitates as a means of establishing reliable use of the TOM.

Applying the TOM

The TOM has now been in use in routine clinical practice, audit, and research as an outcome indicator for over 18 years. It has been translated into Swedish, and has been adapted for use in Australia (Perry, Morris, Unsworth, Duckett, Skeat, Dodd, et al., 2004). While therapists have used the TOM in routine clinical practice, it has also been used as an outcome indicator in research studies (Hammerton, 2004; Marshall, 2004; Parker, Oliver, Pennington, Bond, Jagger, Enderby, et al., 2009) and in clinical audit (Hunt & Slater, 1999) as part of clinical governance.

Enderby and her research team have conducted a number of studies using the TOM as an outcome indicator. These studies have involved working with rehabilitation professionals, including SLPs, physiotherapists, occupational therapists, nursing staff, and intermediate care teams, and have investigated the outcomes associated with different client groups and the similarities and differences in outcomes of care provided by different services (Enderby, et al., 1998; Enderby & John, 1999; John, Hughes, Enderby, 2002; Enderby, Hughes, John, & Petheram, 2003; John et al., 2000; John, Enderby, Hughes, & Petheram, 2001; John, Enderby, & Hughes, 2005a,b; Nancarrow, Enderby, Moran, Dixon, Parker, Bradburn, et al., 2009; Parker et al., 2009; Roulstone, John, Hughes, & Enderby, 2004). The use of the TOM in studies to benchmark outcomes in SLP and other rehabilitation services has utilised internal and external benchmarking strategies to produce outcome data for individual services and to provide overall benchmarks for comparative purposes. Such benchmarking has been useful as a means of assessing clinical practice, allowing comparisons within and between specified services by providing a baseline and permitting repeated measures (Bullivant & Roberts, 1997; MacDonald & Tanner, 1998). It has enabled clinicians to learn from variation and has contributed to overall knowledge. Furthermore, it has demonstrated its utility in the absence of other information on, for example, what works best and for whom, and when relationships between process and outcomes of care are not well understood. When TOM data is combined with structure and process data, and any other relevant qualitative information is added, a body of knowledge is created which can inform practice. Internal benchmarking can thus allow review of a service's

own information, identify areas for change, and compare clinical information with guidelines, standards, and research evidence. Repeating the process allows the effectiveness of changes to be monitored, and, once a service has completed its own quality cycle using external benchmarking, comparisons can be made with other, matched, services to compare outcomes and to ascertain to what extent differences identified can inform change for better outcomes.

In the TOM benchmarking study (Enderby & John, 1999; John, 2001) undertaken between 1998–1999, therapists from 11 different SLP, physiotherapy, and occupational therapy services were trained on the TOM and were reliability checked, after which TOM data was collected on more than 10 000 consecutive cases over an 18-month period.

The study itself sought to answer specific questions, as follows (Enderby et al., 2006; John 2001):

- Do individuals have an equal opportunity of accessing therapy regardless of where they are treated?
- Does therapy produce similar changes regardless of where they are treated?
- Do individuals have similar profiles of abilities and/or disabilities on discharge regardless of where they are treated?
- Is there a difference in the quantity and duration of treatment and is this associated with the amount of change?

The TOM data was analysed using the TOM data analysis program developed by Petheram at Frenchay (described in Enderby et al., 2006) to provide an average rating on each dimension, establishing the benchmark by client group for all cases. In line with the benchmark cycle (MacDonald & Tanner, 1998), individual sites had their own average rating internal benchmark, which could be compared with the overall benchmark for all cases and against other individual sites. At one level, therapists could look at their own outcome results for individual client groups, compare these outcome results with those of therapists in the same service (internal benchmarking), and then compare them with those of other services and with the aggregated data for a specific client group (external benchmarking). The data generated by the TOM program was able to show typical start ratings, change, final ratings, contacts, duration of treatment, and reasons for discharge. Therapists from each service were able to interpret the meaning of the data and to reflect on the information gained, allowing them to decide which aspects of service delivery they wanted to change.

The results of the analysis varied according to the client group. The results arising from the TOM data analysis can be illustrated using data collected on developmental language disorders, analysed to answer the specific questions posed. The data of interest was in respect of: the individual's age on entry to treatment; the severity rating of new cases accessing

treatment; the change on TOM during treatment; the severity of cases at the end of treatment; variation in contacts and duration; and on the reason for discharge. The benchmarking data obtained from the TOM analysis highlighted differences across the participating SLP services. For example, variation was noted in the age when children with a developmental language difficulty were admitted to therapy, and while there was a wide age span at admission to treatment, the majority of cases were at the 4-year age level. There was a statistically significant difference in the admission age in those children diagnosed with a language delay (mean 4.5 years) vs a language disorder (mean 5.1 years). Those children with a language disorder had a more severe TOM admission rating (rated severe) than those children with a language delay (rated moderate). The dimensions of impairment, activity, and participation were the dimensions that had the most positive change on TOM for children with a language delay. The children with a language disorder had similar changes across the dimensions of impairment, activity, and well-being, but had a significant difference on participation, showing more negative changes or sustainment of their difficulties compared with children with a language delay. Across the SLP services, 47% of cases were still in treatment, compared with 53% of cases that had completed treatment by the end of the 18-month study. On final TOM ratings, variation in final rating was evident between cases with a delay vs a disorder, within and across the four services that provided cases with a final rating. There was also a trend for cases with a language disorder to finish treatment at a lower point on the TOM than those with a language delay. Furthermore, those children with a language disorder who completed treatment had more contacts over a longer duration than those with a language delay (benchmark 9.2 contacts all cases, eight contacts language delay, 15.6 contacts language disorder, benchmark 12.9 months duration, 12 months language delay, 16.1 months language disorder). There was a wide variation across the services in the intensity of treatment, as seen with the contacts and duration of intervention. There was a highly significant difference between language delay and language disorder on reasons for discharge, both on the benchmark and across the services. Those cases with a delay had higher self-discharges, and more variation in the reason for discharge, than cases with a language disorder. Overall, the majority of children with two TOM ratings had recorded a general benefit, particularly for those cases that completed their period of intervention. Furthermore, the trend was for a higher percentage of cases with a single disorder to make positive changes than those with a double disorder. The results of this study indicated that, despite variation in the outcomes of care across the participating SLP services, SLP was generally effective.

The reasons for obtaining these outcomes were not known, as information linking structures, pro-

cesses, and outcomes was not collected by the researchers as part of the study. Questions also arose from the study relating to structure and process, the timeliness of interventions, the identification of the correct point to treat, the intensity of treatment, and how much intervention, and by whom, makes a difference. However, SLPs in the participating services were able to consider the outcomes data provided for their service. This allowed them the opportunity to reflect on the similarities and variation in outcomes across the SLP services on the developmental child language client group. The participating services were able to use the TOM data program to generate different data and, by comparing the outcome results (including differences in entry points to the services, the intensity of SLP provided, the change effected and the reasons for discharge), to decide whether any areas of service needed to be addressed. Having established a baseline, the SLPs could build a picture of performance over a period of time, thus informing change for quality improvement and to facilitate the delivery of better care. By identifying an area to change, a service can therefore plan what to change, can collect and analyse the data, add information to the data, and adapt the services provided. As an example, one service noted that newly-qualified SLPs provided more treatment over a longer period than more experienced SLPs. That particular service was able to change its support system in order to provide recently-qualified staff with support in decision-making that concerned the intensity of treatment and appropriate time for discharge. This in turn resulted in better identification of discharge points for planned discharge. In another service, the timing of referrals of children was linked to the timing of pre-school checks by health visitors or referrals at school entry, leading to late referral of children with language disorders. Service changes included putting SLP communication training into nurseries, facilitating language development and earlier identification of children with speech, language, and communication difficulties. In routine clinical practice, experience gained in using the TOM confirms that data collection systems need to allow linkages to be made between structure, process, and outcome data in such a way that does not impose too onerous a burden on SLPs. The TOM can be used in conjunction with other measures. This may be desirable where specific information is required, such as change in phonology vs receptive or expressive language. Studies such as that by Boyle, McCartney, O'Hare, and Forbes (2009) show the value of investigating change in different areas of language. The East Kent Outcomes System (Johnson & Elias, 2002) is one example of a system that incorporates the collection of data on the health benefit of intervention, process, and goal data, with the recording of TOM ratings. In rehabilitation teams, multiple outcome measures can help to inform on different aspects of care, and focus on

specific areas which are key outcome indicators for a client group. Whilst the TOM can be used alongside other outcome measures, it can also be used with tools that aim to capture the reason to treat such as the Malcomess Care Aims or, in intermediate care, eight levels of care developed by Enderby and Stevenson (2000).

Main contribution

The TOM was developed to help SLPs gather information and to reflect on their own performance, thereby contributing to the improvement, year-on-year, in the quality of the service provided. By adding to the overall body of knowledge via the TOM, SLPs are able to consider equity of care, ensuring that individuals have access to SLP services and can obtain intervention at the right time, in the right intensity, and which addresses their needs holistically.

In the context of the UK, characterized by the reorganization of services and pressure on resources, this presents obvious challenges. In particular, those who are commissioning services need evidence of services that are both of high quality and productive within the available resources. Where extra funding is requested, they need to be shown that the money is being well spent. The onus will therefore remain on SLP services to provide evidence of returns on investment in the provision of services for people with speech, language, communication, and swallowing needs. Part of this evidence will derive from the ability to demonstrate improvements in outcomes for individuals, families, and society, an area to which the TOM, together with other measures, makes a significant contribution.

While SLPs may recognize the need to use an outcome tool, a number of issues are intrinsic to the introduction of any measure. For example, SLPs need to believe that the outcome measure they adopt is valid, reliable, and will give them relevant information in a form that is readily understood. There may be concerns about the sensitivity of the measure, for example; TOM does not detect small changes rather clinically significant changes. Equally, there may be concerns about how the outcome information will be applied, including its potential use to justify reductions or cuts to services. Furthermore, in addition to the steps required in order to put an outcome measure such as the TOM into place (including training and reliability checks), SLP services need to establish a system of data entry that addresses structure (staff, skillmix, location), process (intervention applied, intensity of treatment), and outcome measures that together facilitate the generation of reports. To date in the UK, the dearth of resources and systems has generally precluded the production by SLPs of such reports.

There is a need for more research on what works best for whom, and knowledge is still limited on the inter-relationship between process, outcome, and other intervening variables including the provision of

SLP, individual characteristics, and environment. In a recent review of the literature on the effectiveness of SLP with specific client groups conducted by Pam Enderby and her research team (Royal College of Speech and Language Therapists membership website) there was found to be a paucity of good quality research studies. As seen, the use of reliable, valid, and acceptable measures assists clinicians in assessing the effectiveness of interventions and in making comparisons across studies. The TOM has made a contribution to SLP research studies, via the benchmark study described, in the triangulation of information as in the intermediate care studies (Nancarrow et al., 2009; Parker et al., 2009), and in comparing outcomes for individual cases across different dimensions, specifically voice, speech, and swallowing (Radford, et al., 2004).

Underlining the contribution made by outcomes, other countries have created national benchmarking databases both to facilitate the sharing of comparative data and for outcome measures. In the US, for example, a Task Force on Treatment Outcomes and Cost Effectiveness (Frattali, 1998) has been established. Similarly, the American Speech and Hearing Association (ASHA) operates a database to collect and aggregate data on outcomes using specific outcome measures, including the Functional Information Measure (FIM) and Functional Assessment Measure (FAM) (Uniform Data System for Medical Rehabilitation; Turner-Stokes, Nyein, Turner-stokes, Gatehouse, 1999). In Australia, the Commonwealth has established national datasets for Health Care, and the Australian Outcome Measures (AusTOM) (Perry et al., 2004) has been adopted as one of the standard outcome measures. In the UK, however, although there are a number of Government datasets including the National Clinical and Health Outcomes Knowledge Base, there currently exists no national dataset of outcomes measures for SLPs; clinicians are therefore able to choose multiple measures to suit the needs of their service.

Conclusion

In SLP, multiple outcome measures will continue to be appropriate and will vary according to the information needed to show effectiveness and efficiency in delivering specialist and targeted services. Measures are needed to demonstrate the prevention of communication or swallowing difficulties, or in demonstrating reducing demands on services as a result of training and education as well as outcomes of specific interventions. With SLPs working across health, education, social, and voluntary agencies, it is important that these agencies understand the role of SLP and how SLP services are able to contribute (RCSLT, 2009). A shared understanding of each other's contribution will therefore become increasingly necessary, with outcomes encompassing health and social gains and benefits, both for individuals and their families and carers. As

illustrated by this article, the TOM continues to make a meaningful contribution to the assessment of those gains and benefits. It thus stands as another example of the contribution made by Pam Enderby to the continuing improvement in the quality of SLP service delivery.

References

- Bauman, M. K. (1991). The importance of outcome measurement in quality assurance. *Holistic Nursing Practitioner*, 5, 8–13.
- Boyle, J., McCartney, E., O'Hare, A., & Forbes, J. (2009). Direct versus indirect and individual versus group modes of language therapy for children with primary language impairment: Principle outcomes from a randomised controlled trial and economic evaluation. *International Journal of Language and Communication Disorders*, 44, 826–846.
- Bullivant, J., & Roberts, A. (1997). Service benchmarking: To choose to improve. *Health Service Manager* (Special Report no. 14). Kingston upon Thames: Croner Publications.
- Department of Health. (1996). *A service with ambitions*. HMSO: London.
- Department of Health Primary and Community Team and Primary Care Contracting. (2009) *Primary and community care services: improving quality in primary care*. Available online at: www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_106594, accessed 15 May 2010.
- Enderby, P. (1992). Outcome measures in speech therapy: Impairment, disability, handicap, and distress. *Health Trends*, 24, 61–64.
- Enderby, P., & John, A. (1997). *Therapy outcome measures: Speech-language pathology*. San Diego, CA: Singular Publishing.
- Enderby, P., & John, A. (1999). Therapy outcome measures in speech and language therapy: Comparing performance between different providers. *International Journal of Language and Communication Disorders*, 34, 417–429.
- Enderby, P., & Stevenson J. (2000). What is intermediate care? Looking at needs. *Managing Community Care*, 8, 35–40.
- Enderby, P., Hughes, A., John, A., & Petheram, B. (2003). Using benchmark data for assessing performance in occupational therapy. *British Journal of Clinical Governance*, 8, 290–295.
- Enderby, P., John, A., Hughes, A., & Petheram, B. (2000). Benchmarking in rehabilitation: Comparing physiotherapy services. *British Journal of Clinical Governance*, 5, 86–92.
- Enderby, P., John, A., & Petheram, B. (1998). *Therapy outcome measures: Physiotherapy, occupational therapy, rehabilitation nursing*. San Diego, CA: Singular Publishing.
- Enderby, P., John, A., & Petheram, B. (2006). *Therapy outcome measures for the rehabilitation professions: Speech and language therapy; physiotherapy; occupational therapy; rehabilitation nursing; hearing therapists* (2nd ed.). UK: John Wiley & Sons.
- Frattali, C. (1998). *Measuring outcomes in speech-language pathology*. New York: Stuttgart, Thieme.
- Hammerton, J. (2004). An investigation into the influence of age on recovery from stroke with community rehabilitation. *Unpublished PhD Thesis*. University of Sheffield, Sheffield, UK.
- Hesketh, A., Long, A., Patchick, E., Lee, J., & Bowena, A. (2008). The reliability of rating conversation as a measure of functional communication following stroke. *Aphasiology*, 22, 970–984.
- Hunt, J., & Slater, A. (1999). From start to outcome - and beyond. *Speech and Language Therapy in Practice*, 4–6.
- John, A. (2001). Therapy outcome measures for benchmarking in speech and language therapy. *Unpublished PhD Thesis*. University of Sheffield, Sheffield, UK.
- John, A., & Enderby, P. (2000). Reliability of speech and language therapists using therapy outcome measures. *International Journal of Language and Communication Disorders*, 35, 287–302.
- John, A., Enderby, P., & Hughes, A. (2005a). Benchmarking outcomes in dysphasia using the therapy outcome measure. *Aphasiology*, 19, 165–178.
- John, A., Enderby, P., & Hughes, A. (2005b). Comparing outcomes of voice therapy: A benchmarking study using the therapy outcome measure. *Journal of Voice*, 19, 114–123.
- John, A., Enderby, P., Hughes, A., & Petheram, B. (2001). Benchmarking can facilitate the sharing of information on outcomes of care. *International Journal of Language and Communication Disorders*, 36(Suppl), 385–390.
- John, A., Hughes, A., & Enderby, P. (2002). Establishing clinician reliability using the therapy outcome measure for the purpose of benchmarking services. *Advances in Speech-Language Pathology*, 4, 79–87.
- Johnson, M., & Elias, A. (2002). *East Kent Outcome System for Speech and Language Therapy*. East Kent: East Kent Coastal Primary Care Trust.
- MacDonald, J., & Tanner, S. (1998). *Understanding benchmarking*. London: Hodder & Stoughton.
- Marshall, M. (2004). *A study to elicit the core components of stroke rehabilitation and the subsequent development of a taxonomy of the therapy process*. Unpublished PhD Thesis. Sheffield University, Sheffield, UK.
- Nancarrow, S., Enderby, P. M., Moran, A., Dixon, S., Parker, S., Bradburn, M., et al. (2009). *The impact of workforce flexibility on the cost and outcomes of older people's services*. National Institute for Health Research: Service Delivery and Organisation Programme. Project Number: 08/1519/95. University of Sheffield: Sheffield.
- National Clinical and Health Outcomes Knowledge Base. Available online at: www.nchod.nhs.uk, accessed 14 May 2010.
- National Health Service Read Codes. Available online at: <http://www.connectingforhealth.nhs.uk/systemsandservices/data/readcodes>, accessed 15 May 2010.
- National Health Service Institute for Innovation and Improvement World Class Commissioning (2009). Available online at: www.institute.nhs.uk/commissioning, accessed 1 June 2009.
- National Health Service Institute for Innovation and Improvement. Available online at: www.institute.nhs.uk/qualitytools, accessed 15 May 2010.
- Parker, S., Oliver, P., Pennington, M., Bond, J., Jagger, C., Enderby, P.M., et al. (2009). Rehabilitation of older patients: Day hospital compared with rehabilitation at home. A randomised controlled trial. *Health Technology Assessment UK*, 13, 168.
- Perry, A., Morris, M., Unsworth, C., Duckett, S., Skeat, J., Dodd, K., et al. (2004). Therapy outcome measures for allied health practitioners in Australia: The AusTOMs. *International Journal for Quality in Health Care*, 16, 285–291.
- Radford, K., Woods, H., Lowe, D., & Rogers, S. (2004). A UK multi-centre pilot study of speech and swallowing outcomes following head and neck cancer. *Clinical Otolaryngology and Allied Sciences*, 29, 376–381.
- Roulstone, S., John, A., Hughes, A., & Enderby, P. (2004). Assessing the validity of the therapy outcome measure for pre-school children with delayed speech and language. *Advances in Speech-Language Pathology*, 6, 230–236.
- Royal College of Speech and Language Therapist (2009). Available online at: http://www.rcslt.org/speech_and_language_therapy/intro/resource_manual_for_commissioning_and_planning_services, accessed 1 January 2010.
- Stevenson, J., & Spencer, L. (2002). *Developing intermediate care: A guide for health and social service professionals*. London: King's Fund.
- Swage, T. (2000). *Clinical governance in healthcare practice*. Oxford: Butterworth & Heinemann.
- Turner-Stokes, L., Nyein, K., Turner-Stokes, T., & Gatehouse, C. (1999). The UK FIM+FAM: Development and evaluation. Functional Assessment Measure. *Clinical Rehabilitation*, 13, 277–287.
- World Health Organization. (1980). *International Classification of Impairments, Disabilities and Handicaps (ICIDH)*. Geneva: WHO.
- World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO.
- Uniform Data System for Medical Rehabilitation. Available online from: http://www.udsmr.org/WebModules/FIM/Fim_About.aspx, accessed January 2010.