**7. Document analysis**

**Jane O’Connor**

**Introduction**

This chapter provides a discussion of document analysis as a method for social and educational enquiry. It begins with an overview of the definition and use of this approach as part of the larger notion of ‘discourse analysis’, and goes on to explain how it was employed by the author in a research project which explored how gifted and talented children are portrayed in the British press. Processes around data collection and analysis are carefully considered and explained, interpretation of findings are discussed, and the final section provides guidance on how to decide if document analysis might be a suitable method to use in your research project.

**Discourse and document analysis**

Document analysis is part of a raft of methodological approaches which come under the umbrella term of ‘discourse analysis’. The aim of discourse analysis is to investigate the social meanings inherent in spoken language, images and text. Methods range from micro-analysis of every utterance, pause and intonation in spoken language, using the specific techniques of conversation analysis (see Sidnell, 2010), to document analysis itself, which focuses on illuminating the themes and ideologies which give meaning to pieces of writing, such as newspaper articles, company websites or policy documents. The fundamental theoretical concept of all discourse analysis, however, is that the surface level of language, in whatever way it is presented, is the ‘tip of the iceberg’ and that this hides a vast array of socially constructed and culturally shared information and knowledge which give the spoken language, images or text meaning to the audience to which it is addressed.

By giving us strategies for unpicking the underlying, socially constructed knowledge which informs the content and form of the language we use to communicate, both formally and informally, discourse analysis allows us to analyse texts, images and speech acts critically, in order to challenge the assumed, often biased or unjust assumptions and the ‘common-sense’, shared ideologies which inform them. In this sense the key questions to ask when approaching data from any discourse-analysis perspective are these:

* Why are these words or images being used and not others?
* What is being assumed or implied about the subject in this discourse material?
* What shared cultural or social knowledge is allowing me to understand and make meaning from it?
* How are individuals or groups of people being positioned in this material?

Discourse analysis, and specifically document analysis, have often been usefully employed as tools for highlighting and uncovering social issues. For example, van Dijk (2016) reviewed use of the discourse-analysis approach in studying racism, including investigating the role of the Dutch media in the reproduction of racism; analysis of everyday conversations and storytelling about immigrants; and the representation of immigrants and third-world peoples in Dutch social-science textbooks. Baker (2014) used document analysis to explore gender and sexuality in a case study of a British newspaper article about the death of the boy-band singer, Stephen Gately, in 2009. Baker’s analysis, both of the newspaper report and of the online responses to it, critically examined features of sexual-identity discourse in the British media at that time. Fealy and McNamara (2009) used a similar approach in their analysis of two daily newspapers over a key period in Irish politics, when in 2008 the Irish Government withdrew automatic entitlement to free medical care for people aged 70 and over. They identified in these media particular ways of naming and referencing older people and also revealed distinct and discernible constructions of ageing and of age identities. Their research demonstrated the ways in which older people are positioned in discourses and how these resultant identities have consequences for their behaviour and for the way that other individuals and society behave towards them.

These examples show how discourse and document analysis can be used for highlighting dominant, often unjust discourses which underpin governmental policy, media and other material around important issues in education, health and social equality. Such discourses are frequently invisible or well hidden within rhetoric and euphemistic language. The following section outlines how document analysis was used in a piece of education-studies research (O’Connor, 2012), which explored how the concept of childhood ‘giftedness’ was constructed and written about in the British press in 2006-2008, and how this related to the educational practice of labelling certain children as ‘gifted’ or ‘talented’ in English schools, which was current policy at that time.

**Using document analysis in a research project: Is it good to be gifted?**

This research project was carried out in the context of a new educational policy initiative in England in the 2000s, which required schools to identify up to 10 per cent of children as either gifted academically or talented in sports or the arts, then to provide them with specialized support and opportunities to extend and enrich their learning (DCSF, 2009). It is worth noting here that this distinction between ‘gifted’ and ‘talented’, which was made in Government documents, tended not to be used more widely. Indeed, in the media the term, ‘gifted’, was and still is often applied to those who excel in any area.

As a childhood-studies expert, I had previously investigated the stigmatizing experience that being labelled as extraordinary in any field could have on children (O’Connor, 2008).

Therefore, in this subsequent research study I was concerned about the way in which being labelled as gifted or talented may be perceived by the general public and about the unintended adverse outcomes that such a practice may have on the well-being of a potentially large number of children. This was reinforced when I reviewed associated research in this field and discovered that socially negative perceptions around exceptionally able children were a particular problem in other countries, which had already established a similar educational policy around classifying children as gifted or talented. For example, Geake and Gross (2008) conducted research with trainee teachers in Australia and found that negative attitudes towards students labelled in this way were prevalent.

I reflected on how best to explore this issue in England and decided that it would be interesting and useful to look from a social perspective at how the media conceptualizes and presents children by employing such terms, as this is one way in which any associated stigma might be reinforced and circulated. That is not to say that other approaches, such as a study of individual experiences or a survey of parental or practitioner responses to this policy, would not also have been important ways to investigate this topic. With the document-analysis approach, however, I wanted to achieve a different, wider way of seeing the issue, one which was not situated in specific schools or families or professional groups, but one which recognized the practice of labelling as a powerful social act, part of a broader process in which we, as a society, construct childhood in certain bounded ways, and which might show how we understand and use related categories, such as giftedness. The concept of childhood as being socially constructed underpinned the research theoretically and allowed me to explore giftedness as an ‘invented’, rather than a‘natural’ category, which in turn allowed me to critically examine its use in educational policy in England.

[BOX BEGINS]

**The social construction of childhood**

*Within the dominant discourse that is recognized as characterizing contemporary Western constructs, childhood is understood as being separate from adulthood, and children as less knowledgeable and skilful than adults (Hockey and James, 1993). As Archard (1993:30) states: ‘Childhood is defined as that which lacks the capabilities, skills and powers of adulthood.’ This clear distinction not only legitimizes unequal power relations but also validates the long period of compulsory education that characterizes the increasingly global definition of how children should spend their time. In this sense, the child who is academically gifted upsets these normalized relations and can be seen to be transgressing the boundaries between childhood and adulthood which have been so carefully constructed in Western society over the last 300 years (Ariès, 1996). Such children also challenge the developmental ‘ages and stages’ approach to organizing education systems, based on the work of Piaget, that structures the vast majority of schooling across Europe and the Western world (Bentham, 2002).*

Find and read a newspaper article about a child or children. Then use the ideas presented in the review above to inform your responses to the following questions:

1. What assumptions or shared ideologies about childhood are evident in the newspaper report you have chosen? For example, it might assume that readers agree that childhood is a time of innocence, or that children should be protected, or that teenagers are dangerous.
2. What specific words (for example, ‘angels’ or ‘hoodies’) are used to describe children in the article? How might these labels influence or reinforce readers’ behaviour and attitudes towards children and young people?
3. Think about the wider context of the newspaper article you have found. How might the same story have been written 20 years ago? One hundred years ago? Or in a different country or culture?

[BOX ENDS]

**Data collection**

The first step in my research was to identify and collate a data set with which I would explore my chosen issue. I decided to use articles from a range of British broadsheet and tabloid newspapers from a two-year period between March 2006 to December 2008. The data was collected via the free (at that time) newspaper database, Proquest, using the search terms ‘child prodigy’, ‘gifted child’ and ‘talented child’. In all, 187 stories were found, many being about the same children. Of these stories, 17 related to academic giftedness, 15 to high ability in sport and 13 to high ability in music: 45 relevant reports in total. Twenty-eight of these were about boys, 17 were about girls. The stories were about British children and other high-profile, exceptional children from across the world.

One advantage of using document analysis is the convenience of being able to collect data from publicly available, web-based databases, which saves both time and resources at this early stage of the research. It is important to define your search terms, also to be flexible in adjusting them to obtain the stories which cover the area you are interested in, as you become more familiar with the kind of language used around the topic. For example, in my initial search I did not use the term, ‘child prodigy’, as I perceived it as too unusual and rarely used. But when I did include it, many extra articles came up, as this is actually a common way for British journalists to refer to any child of above-average ability, especially in tabloid newspapers.

There is also the opportunity to be flexible in relation to your search period. For example, if a search over a one-year period does not yield many results, then it can be extended to two years, and so on, until you have collected a critical mass of stories. It is important, however, to provide some rationale for the range of material and the time period you are using. For example, in my research I purposefully chose a period of time when Government policy was developing in the area of gifted-and-talented provision.

There is no definitive way of knowing when you have enough data, as the number of documents available on different topics varies immensely. It also depends on your research design and aims. For example, if you are investigating formulation of a government policy, then the main policy document alone could be your data set. However, where there is a large amount of potential data, I tend to use the grounded-theory concept of ‘saturation’ (Charmaz, 2014) to know when I have collected enough documents. Saturation refers to the point when the same themes, narratives and language keep coming up again and again, and nothing novel seems to be contributed to the data by adding more material. That is generally the point when I can begin unpicking the deeper discourses that are informing the stories and start to identify the underlying themes and assumptions that are supporting and giving meaning to the texts.

So it is important that you are both flexible and systematic in your data collection and are able to explain clearly the provenance and scope of your final data set. This will lend trustworthiness and validity to your research and give you the basis to address, in the methodology or other section of your written project, two important questions: Have I gathered enough data to make a convincing case that my findings are trustworthy? Have I gathered it in a transparent way that could be replicated by another researcher? It is important to show that you did not ‘cherry-pick’ only the articles that fitted your assumptions about the research questions and discard any that did not.

Another positive aspect of undertaking document analysis is that the absence of human subjects means there are fewer ethical considerations to take into account. It is not necessary to obtain permission to use databases which contain public-domain material that has already been published, such as newspaper articles (although some publications charge a fee to access their archives). There is also no need normally to obtain consent from the individual journalists or subjects of the stories, although all sources do need to be referenced in your research paper, just like academic texts. Document analysis involving personal information (such as diaries) or confidential material (such as health and education records) would, however, necessitate permission for access from the individuals involved and stringent ethical approval on several issues, including informed consent, anonymity and storage of data (BERA, 2018). If in doubt about whether to seek permission to scrutinize a particular document in your research (as might be the case, for instance, with material coming from a school), you may wish simply to inform those responsible for the document about what you are doing, and keep details of its origin confidential in your written project report.

[BOX BEGINS]

**Planning your research**

1. What kind of documents could you analyse in order to investigate your research topic?
2. How would you access these documents? What challenges, if any, might be involved?
3. What ethical issues would you need to consider? What steps would you need to take to address these issues?

[BOX ENDS]

**Analysis and findings**

To analyse and interpret the textual data collected for the gifted study, a qualitative approach was taken. This involved carefully reading all the newspaper articles in the sample and coding them according to categories which related to how the children were described. This was done by printing out the stories and using coloured highlighter pens to identify extracts or quotes with similar themes (this process of ‘coding’ can also be undertaken using specially designed software, such as NVivo). Common themes and patterns were identified in the discourses and terminology was noted. For example, articles using the term, ‘boffin’, were found to be generally less positive about childhood giftedness than those using the term, ‘genius’. The portrayal of the three types of giftedness – academic, sport, music – were also compared in order to highlight any differences in the way each was represented. Finally, the texts were analysed in relation to dominant social constructions of childhood as a time of learning and of children being less knowledgeable and skilful than adults, as shown earlier. This was done in order to identify the potential stigmatizing effects of labelling children as gifted.

Themes identified in the analysis included ‘gifted children as deviant’, ‘having a normal or abnormal childhood’, and ‘attribution of ability’ – for example, was the child described as being born with a gift or working hard to achieve it? These formed the structure for presentation of the analysis of findings in the final journal article (O’Connor, 2012). It is important to note that document analysis can also include a quantitative approach to texts, whereby, for example, instances of the use of certain words or phrases pertinent to the research topic are counted and their frequencies analysed and discussed. This detailed scrutiny of language and word frequency was employed in Fealy and McNamara’s (2009) examination of newspaper reactions to changes to medical care in Ireland, described earlier.

From my data analysis, I found that newspaper stories about academically gifted children tended to be constructed to highlight differences between the ‘normal’ and the ‘gifted’ child, and in doing so, reinforced and naturalized an idealized image of the child as occupying a temporal, cognitive space of pre-adult understanding and knowing. The construction of the academically gifted child as a passive pawn of adult manipulation was a common theme in stories in which the prodigy had encountered misfortune or disaster, reflecting the way in which the press standardize a certain idea of the ‘proper child’ as one who is not pushed beyond an intellectual stage, socially agreed as appropriate for their years.

The difficulty of striking a balance between being ‘pushy’ and being ‘encouraging’ was particularly pronounced in the stories about the parents of academic prodigies, with many commenting on the ‘problem’ of their child’s gift and the burden of responsibility for organizing the best provision for his or her needs. It was also evident from the data that the dangers of academically precocious children being seen as strange were compounded by the tendency of newspaper stories to depict them as oddly different from ‘normal’ children.

Another key finding was that children who were identified as gifted in music or sport did not appear to be subject to the same negativity in the newspaper stories as those who were said to be gifted academically. Indeed, the stories seemed much less hostile towards young musicians and sports stars than towards their academic counterparts, and a more admiring, proud tone often prevailed when they were described. In the article I theorize that this difference in tone can be explained from a ‘social-constructionist’ perspective, with children who are exceptional in the fields of music and sport ‘constructed’ in the newspapers in a largely different way from the academically gifted. This variation seems to emanate from different paradigms of giftedness – children gifted musically or athletically are seen as manifestations of different aspects of childhood ‘atypicality’ to those showing academic giftedness, and those aspects have more positive connotations in Western culture because of the wider discourses that inform them.

**Discussion**

It is important to acknowledge the limitations of the document research reported in this chapter, in terms of its relatively small sample of stories and the subjective interpretation of selected data. For these reasons, it was not possible for findings to be generalized to other cultural and social contexts, as they described only how exceptionally able children were being portrayed in the British press at a specific point in time. The value of the study was derived from the way in which it demonstrated the socially constructed nature of childhood giftedness and how public responses to such children appeared to be moulded and constrained by wider social and cultural, normative understandings around childhood more generally. This was only possible through adopting a document-analysis approach to the data. However, even a small-scale study such as this succeeded in providing evidence of differences in social and cultural constructions of these children, and formed part of a critical academic response to the practice of labelling them as ‘gifted’ or ‘talented’ in English schools at that time.

As has been seen, my analysis of documents showed that academically gifted children were subjected to the majority of negative attitudes, stigma and cautionary tales in the newspaper stories. They were portrayed as being outside the ‘normal’ parameters of childhood ability, as ‘swots’ or ‘geeks’, for example, sometimes with the additional assumption that their parents were somewhat deviant as well. Relating this finding to dominant discourses on childhood, it seemed that the academically gifted child transgressed the boundaries between adulthood and childhood by ‘knowing too much, too soon’, and also challenged traditional educational structures regarding incremental exposure to appropriate knowledge at certain ages, as decided by adults.

I extended this thinking further, again by combining findings from the document data with wider social discourses, and again relating them in particular to views on academically gifted children. I concluded that these children can be viewed as being non-compliant and subversive of accepted norms of the teacher-student, or more general adult-child binaries. They also challenge the ideal of childhood innocence by being ‘precocious’ in their intellect, an unwelcome trait in much of Western society, where children are more often valued for their association with naivety and vulnerability than for their accelerated pace of academic learning. The historical associations between premature sexuality and the overly advanced child still appeared to inform the general distaste for the precocious child (or ‘brat’), which was – and probably still is – evident, sometimes quite overtly, in media stories about intellectually able children. The implication from the study was that due to contemporary dominant ideologies of childhood in English society, children labelled as exceptionally able in academic fields could be perceived as challenging adult authority and the received social order around their ‘proper’ place. On a broader level, the findings suggested that it would be socially desirable to move on from such constraining definitions of what is ‘normal’ and what is not, and who is ‘gifted’ and who is not, which may permeate both institutional and cultural discourses on childhood and which ultimately serve to limit the individuality and creativity on which innovation and progress thrive.

[BOX BEGINS]

**Reaching wider conclusions**

1. In your view, to what extent and in what ways was document analysis a useful approach to investigating the research issue in this project?
2. What are your thoughts about the way in which the document analysis was used in the study to reach wider conclusions about education, society and the media?
3. What did the use of document analysis add to the debate that other methods may not have contributed?

[BOX ENDS]

**Conclusion**

Techniques of document analysis are developing all the time, and the quantity and range of potential material to analyse proliferates daily, thanks mainly to the Internet and digital technology. Some researchers use specially designed software and linguistic programs to analyse large data sets of text, a process called ‘corpus linguistic analysis’. Another term, ‘text mining’, is often given to the process of deriving high-quality [information](https://en.wikipedia.org/wiki/Information) from large amounts of online [text](https://en.wikipedia.org/wiki/Plain_text) through the identification of patterns and trends. The overarching aim of such approaches is to turn text into data for analysis (see Ignatow and Mihalcea, 2017). In terms of the wider concept of discourse analysis, other researchers have developed innovative techniques for analysing television dialogue (Bednarek and Caple, 2017) and modern ‘pop culture’ (Werner, 2018), with the aim of critically exploring the meanings behind spoken language and images, as well as written texts.

The document analysis method discussed here demonstrates the availability and accessibility of an extensive range of text material for analysis, opening up the possibility of research into many aspects of education and other social issues. However, there are, of course, limitations in the approach, especially if it is not carried out well. For instance, as has been seen, it is important to report on your data collection in a transparent manner, so that readers of your research understand the processes you went through. It may also be necessary to dig down deeper into issues by combining document analysis with other methods, such as interviews, so that your analysis of the documents is not overly subjective or left open to alternative readings. This also places responsibility on the researcher to gather trustworthy and persuasive evidence, to seek ‘saturation’ of data, to be cautious in the conclusions drawn and to acknowledge the limits of generalizing findings. It is also important to be familiar with the cultural and social context from which the documents to be analysed arise, as well as with the language in which they are written, as most interpretations will depend on an understanding of the wider and deeper meaning of the texts. Indeed, document analysis of material from an unfamiliar culture can be very difficult, and in such cases a different methodological approach, or the help of a culturally situated research partner, would be advised.

The question of whether document analysis is right for your research project comes down to what you want to explore, what data is available, and whether you are interested in social discourses as evident in documents, or in exploring what people think, say and do – aspects which might be better investigated through other methods. If you are a student, doing your investigation as part of a course of study, there may indeed be a requirement for you to engage in human contact, rather than documents, for your research project and you may need to check this beforehand with your supervising tutor. However, even if it is not used as your main method, document analysis can be combined with other methods, thereby triangulating your research design, and the findings it produces can complement those from other types of data collection. If you are intrigued by how the wider culture that we live in influences how we understand the world and interact with each other, and you are interested in the often concealed power of language and discourse, then document analysis could well be a method worth using in your research.

To summarize this chapter:

* Document analysis is a form of discourse analysis that uses written texts as its data source.
* It is a useful method for exploring discourses and ideologies that inform official and cultural material.
* It can be used to critically analyse policy documents and identify hidden social inequality and stereotyping.
* Many documents are freely available online and this can simplify data collection, especially if time for a project is limited.

**Recommended Reading**

Bowen, G.A. (2009) Document analysis as a qualitative research method. *Qualitative Research Journal*,9(2), 27-40.

Targeted at early researchers, this article takes a ‘nuts-and-bolts’ approach to document analysis.

Rapley, T. (2008) *Doing Conversation, Discourse and Document Analysis*. London: Sage.

This book provides an overview of the particular challenges involved in collecting and analysing data, as well as clear guidance for each stage of your project.

Strauss, S. and Feiz, P. (2014) *Discourse Analysis: Putting Our Worlds into Words.* Abingdon: Routledge.

This is another useful resource for understanding the broader techniques and uses of discourse analysis.

**References**

Archard, D. (1993) *Children: Rights and Childhood*. London: Routledge.

Ariès, P. (1996) *Centuries of Childhood*. London: Pimlico.

Baker, P. (2014) Considering context when analysing representations of gender and sexuality: A case study. In: J. Flowerdew (Ed.) *Discourse in Context: Contemporary Applied Linguistics Volume 3*. New York: Bloomsbury.

Bednarek, M. and Caple, H. (2017) *The Discourse of News Values: How News Organizations Create Newsworthiness*. New York: Oxford University Press.

Bentham, S. (2002) *Psychology and Education*. Hove: Routledge.

BERA (2018) *Ethical Guidelines for Educational Research*. Fourth edition. London: British Educational Research Association.

Charmaz, K. (2014) *Constructing Grounded Theory.* Second edition. London: Sage.

DCSF (2009) *Standards Site: Gifted and Talented*. [Online] http://www.standards.dfes.gov.uk/giftedandtalented (accessed 12 January 2010).

Fealy, G. and McNamara, M. (2009) *Constructing Ageing and Age Identity: A Case Study of Newspaper Discourses.* Dublin: National Centre for the Protection of Older People.

Geake, J.G. and Gross, M.U.M. (2008) Teachers’ negative affect toward academically gifted students: An evolutionary psychological study. *Gifted Child Quarterly*, 52(3), 217-231.

Hockey J. and James A. (1993) *Growing Up and Growing Old: Ageing and Dependency in the Life Course.* London: Sage.

Ignatow G. and Mihalcea, R.F. (2017) *Text Mining: A Guidebook for the Social Sciences*. London: Sage.

O’Connor J. (2008) [*The Cultural Significance of the Child Star*](http://www.routledgesociology.com/books/The-Cultural-Significance-of-the-Child-Star-isbn9780415961578). Abingdon: Routledge.

O’Connor J. (2012) Is it good to be gifted? The social construction of the gifted child. *Children and Society*, 26(4), 293-303.

Sidnell, J. (2010) *Conversation Analysis: An Introduction*. Chichester: Wiley-Blackwell.

Van Dijk, T.A. (2016) Discourse analysis of racism. In: J.H. Stanfield (Ed.) *Rethinking Race and Ethnicity in Research Methods*. Abingdon: Routledge.

Werner, V. (2018) *The Language of Pop Culture*. Abingdon: Routledge.