Structural topic modeling as a mixed methods research design: a study on employer size and labor market outcomes for vulnerable groups

Julie Ulstein[1]

Abstract

Obtaining and maintaining steady employment can be challenging for people from vulnerable groups. Previous research has focused on the relationship between employer size and employment outcomes for these groups, but the findings have been inconsistent. To clarify this relationship, the current study uses structural topic modeling, a mixed methods research design, to disclose and explain factors behind the association between employer size and labor market outcomes for people from vulnerable groups. The data consist of qualitative interview transcripts concerning the hiring and inclusion of people from vulnerable groups. These were quantitized and analyzed using structural topic modeling. The goals were to investigate topical content and prevalence according to employer size, to provide a comprehensive guide for model estimation and interpretation, and to highlight the wide applicability of this method in social science research. Model estimation resulted in a model with five topics: training, practicalities of the inclusion processes, recruitment, contexts of inclusion, and work demands. The analysis revealed that topical prevalence differed between employers according to size. Thus, these estimated topics can provide evidence as to why the association between employer size and labor market outcomes for vulnerable groups varies across studies—different employers highlight different aspects of work inclusion. The article further demonstrates the strengths and limitations of using structural topic modeling as a mixed methods research design.

Keywords Labor market outcomes · Mixed methods · Structural topic modeling · Vulnerable groups · Work inclusion

✉ Julie Ulstein
  Julie.Ulstein@Oslomet.no

1   Work Research Institute, Oslo Metropolitan University, Oslo, Norway

Springer

1 Introduction

Sustainable employment has gained momentum in recent years. In this study, it is defined as the extent to which workers are able and willing to acquire and maintain steady employment (McCollum 2012; Van Dam et al. 2017), which is important to secure the sustainability of welfare states (McCollum 2012) and foster social integration (De Witte et al. 2016). Vulnerable groups are sub-populations that might need more humanitarian, social, or financial assistance than other groups, due to intrinsic and/or extrinsic characteristics (Marin-Ferrer et al. 2017). In the context of labor market participation, examples of vulnerable groups include people with disabilities and ill health, older workers, people employed in precarious jobs, ethnic minorities, the long-term unemployed, and other groups that face barriers to acquiring and maintaining steady employment (Van Berkel et al. 2017). In addition to intrinsic and extrinsic individual factors, characteristics of the work context can impact sustainable employment (Marin-Ferrer et al. 2017; McCollum 2012; Van Dam et al. 2017). Consequently, studies have investigated how different employer characteristics can affect labor market outcomes for vulnerable groups.

Part of the literature on employer characteristics argues that the ability and opportunity to hire and retain people from vulnerable groups might vary with employer size (Bacon and Hoque 2021; Nagtegaal et al. 2023). Thus, more knowledge of the association between employer size and labor market outcomes could contribute to the development and implementation of policies designed to increase group-level labor market participation among vulnerable groups. However, the evidence on this association remains inconsistent, as different studies have variably found positive associations, negative associations, and no associations between employer size and labor market outcomes for vulnerable groups (Jansen et al. 2021). Few studies have applied multiple or mixed methods designs when investigating this association (Bento and Kuznetsova 2018; Hyggen and Vedeler 2021; Johnston et al. 2015; Kocman et al. 2018), and as the current paper argues, this presents a lacuna in the literature on labor market outcomes for vulnerable groups. Given the complexity of this topic, mixed methods research (MMR) that integrates qualitative methods for deep learning and quantitative methods for broad inference (Bickman et al. 2009; Tzgkarakis and Kritas 2022) would improve the use of data and reveal factors that can explain the inconsistent evidence for the association between employer size and labor market outcomes for vulnerable groups.

In this article, I use structural topic modeling (STM). STM is a research method that integrates qualitative and quantitative methods and data, for which it can be considered an MMR design (Bickman et al. 2009; Roberts et al. 2019). Topic models result from machine learning techniques in which the underlying semantic structure of text documents can be discovered (Barde and Bainwad 2017; Kherwa and Bansal 2019). A topic can be defined as a mixture of words, each of which has a probability of belonging to that topic (Roberts et al. 2019). A document can be composed of multiple topics, for which topical prevalence can vary. In the current study, STM was applied to qualitative interview transcripts to identify the topical content addressed by employers and whether topical prevalence differed according to employer size. The results will contribute to clarifying the inconsistent findings on employer size and labor market outcomes for vulnerable groups. As such, this study aims to answer the following research questions:

1. What topics are addressed by employers and does topical prevalence differ by employer size?
2. How can STM be used as an MMR design in the analysis of qualitative interviews?

To answer the research questions, transcripts of 49 interviews concerning hiring and including people from vulnerable groups were quantitized using STM (Bickman et al. 2009; Roberts et al. 2019). Quantitizing means transforming the interview transcripts into quantitative entities represented by word frequencies within and across interview transcripts. These data were then implemented in a topic modeling framework and subject to interpretation via cross-reference to the qualitative interviews. This integration of qualitative and quantitative data analysis through topical content and topical prevalence enabled both deep and broad analysis of the data.

This study makes two main contributions. First, it presents new knowledge to the field of sustainable employment for people from vulnerable groups as it investigates whether topical content and topical prevalence can reveal variable associations that better explain the inconsistent findings on the relationship between employer size and labor market outcomes for vulnerable groups. Second, the study contributes to MMR and methodology generally by introducing STM as a new method for integrating qualitative and quantitative methods in the analysis of interview transcripts. To my knowledge, no such study has been conducted to date. STM has great potential in social science research, as many research projects have data material consisting of long texts and relatively small sample sizes. STM can be used to systematize collections of interview transcripts and to investigate whether topical content or prevalence varies with document variables. Consequently, this study provides a detailed guide on how interview transcripts can be preprocessed, how model preparation and estimation can be conducted, and how model output can be interpreted using interview transcripts in terms of topics, top predicted words per topic, and topical prevalence.

The next section reviews the literature on employer size and labor market outcomes for vulnerable groups, while the third section presents the selection criteria, sample information, and interview protocol used for data collection. The fourth section describes the process of conducting STM, from preprocessing and selection of the number of topics to model estimation and analysis. The fifth section presents the results of this process, with interpretations of the topical content and a graphical representation of topical prevalence by employer size. The sixth section discusses the results and contributions to the literature and outlines the limitations of both STM and the study. Finally, the seventh section highlights the potential of STM as an MMR design and suggests avenues for future research.

2 Literature review

To increase labor market participation for people from vulnerable groups, previous research has focused on how factors related to the work context can impact labor market outcomes. Employer size is a contextual factor that has received special attention in the literature, as many have theorized that employer size is associated with different labor market outcomes for vulnerable groups (Bacon and Hoque 2021). If employer size is associated with positive or negative employment outcomes for vulnerable groups, it could constitute a viable target

for policy interventions aimed at increasing labor market participation and contributing to sustainable employment.

Many researchers have predicted a positive association between employer size and labor market outcomes for vulnerable groups (Bacon and Hoque 2021; Beatty et al. 2019; Goss et al. 2000). The predicted positive association has been attributed to the higher degree of formality present in larger organizations, typically through the presence of a department or team dedicated to human resources, or to established policies and practices for diversity and equality, both of which can contribute positively to labor market participation and the integration of people from vulnerable groups into the workplace (Bacon and Hoque 2021; Beatty et al. 2019; Goss et al. 2000). Additionally, larger organizations have more resources available to make accommodations and are less sensitive to economic fluctuations. This can make hiring and integrating people from vulnerable groups less of a risk for larger organizations compared to smaller ones.

At the same time, other researchers have predicted precisely the opposite relationship (Stone and Colella 1996). In their theoretical model of factors that can affect the treatment of people with disabilities in organizations, Stone and Colella (1996) argued that the degree of informality and absence of bureaucratic obstacles in smaller organizations can be positive for both employment outcomes and the treatment of people with disabilities because it facilitates greater flexibility to accommodate employees based on individual needs. This is supported by research demonstrating that small employers provide the benefits of high informality, a high degree of flexibility in work hours and tasks, and greater opportunities for personalized treatment (Harney and Alkhalaf 2021; Storey et al. 2010; Tsai et al. 2007). Based on this, a negative association would be expected between employer size and labor market outcomes for vulnerable groups, as the high degree of flexibility and personalized treatment decreases when employer size increases.

The evidence on the association between employer size and labor market outcomes for vulnerable groups remains inconsistent (Jansen et al. 2021). Some studies have found positive associations between employer size and continued employment, return to work (RTW), and employee retention (Hannerz et al. 2012; Prang et al. 2016; Schneider et al. 2016; van Ooijen et al. 2021). Others have found negative associations between employer size and lower entry to certified absence, early RTW, and continued or sustained employment (Faucett et al. 2000; Holm et al. 2007; Krause et al. 2001; Markussen et al. 2011; Ulstein 2023). At the same time, other studies have found no associations between employer size and RTW, continued employment, or number of employees from vulnerable groups (Bacon and Hoque 2021; Cooper et al. 2013; Høgelund and Holm 2014). One explanation for the inconsistent evidence could be that rather than impacting labor market outcomes directly, employer size represents a proxy for employer knowledge, capability, and motivation to hire and retain people from vulnerable groups (Nagtegaal et al. 2023). The knowledge and organizational structures required for employers to hire and retain people from vulnerable groups can include human resource practices (or the absence thereof), practices related to selection and recruitment, the appropriateness and complexity of work tasks, supportive leadership and work environment, collaborations with external actors, and training and development (Beatty et al. 2019; Hulsegge et al. 2022; Jansen et al. 2021; Kersten et al. 2023; Strindlund et al. 2019). Motivation can include corporate social responsibility concerns and expectations of either economic or competitive (dis)advantages through hiring (Nagtegaal et al. 2023; Van der Aa and Van Berkel 2014). Additional research is necessary

to clarify the relationship between employer size and labor market outcomes for vulnerable groups and determine whether other factors behind this relationship can explain the inconsistency of previous findings.

3 Case selection and data collection

This study was performed using interview transcripts on labor market inclusion of people from vulnerable groups. The goal of the interviews was to gain insight into key factors that contribute to the successful integration of people from vulnerable groups in the workplace. The interviews were conducted between September 2021 and June 2022 by a research team of five members located in the greater Oslo region, Southern Norway, Western Norway, and Central Norway. Following the suggestions of welfare administration professionals, organizations were recruited based on their experiences with hiring people from vulnerable groups. The research team contacted the suggested organizations by phone or email, providing a brief description of the study and what participation would entail. To ensure sufficient variation among the organizations, two characteristics were considered. First, the organizations had to vary in number of employees to capture potential differences between micro (<10), small (≥10 and <50), medium (≥50 and <250), and large (≥250) employers. Second, the organizations had to represent different industries to ensure variation in business activities. Additionally, to ensure adequate levels of experience with hiring people from vulnerable groups, the organizations had to have at least one employee from a vulnerable group on an ordinary contract at the time of the interview. An overview of the sample characteristics is presented in Table 1.

The study encompassed 49 interviews with various participating team members (supervisors, HR managers, colleagues, and employees from vulnerable groups) from 17 organizations. The interview guide contained six overarching themes: (i) general information

Table 1 Sample of organizations

Employer	Size	Industry	Interviews
Bakery	Medium	Manufacturing	7
Cafe	Micro	Accommodation and food services	2
Facility services	Large	Administrative and support services	2
Fireplace installer	Micro	Wholesale and retail trade	2
Gadget store	Large	Wholesale and retail trade	3
Gardener	Micro	Administrative and support services	2
IT company	Small	Information and communication	2
Kindergarten	Small	Health and social services	3
Landscaper	Small	Administrative and support services	4
Logistics company	Medium	Transportation and storage	2
Machinery manufacturer	Micro	Manufacturing	1
Masonry company	Medium	Construction	1
Nursing home	Medium	Health and social services	2
Supermarket 1	Small	Wholesale and retail trade	4
Supermarket 2	Small	Wholesale and retail trade	4
Supermarket 3	Small	Wholesale and retail trade	3
Technology business	Large	Manufacturing	5
N	17		49

about the organization; (ii) motivations for and experience with hiring individuals from vulnerable groups; (iii) specific examples of employees belonging to vulnerable groups; (iv) the internal processes involved in hiring, training, and accommodating employees; (v) support from and cooperation with external organizations, such as Norwegian Labor and Welfare Administration (Norges Arbeids- og Velferdsetat; NAV); and (vi) financial support granted to the organization, such as wage subsidies. The interview guide was piloted in interviews with representatives from two organizations reporting dissimilar experiences with hiring people from vulnerable groups, after which the interview guide was adapted. All interviews were recorded and transcribed verbatim.

4 Structural topic modeling

STM is a type of unsupervised topic model, which refers to a method in which machine learning algorithms automatically identify concepts through the clustering of words, groups of words, or texts (Macanovic 2022; Törnberg and Törnberg 2016). There are several unsupervised topic modeling methods available, the most common of which are Latent Dirichlet Allocation (LDA), correlated topic modeling (CTM), and STM (Blei and Lafferty 2007; Blei et al. 2003; Roberts et al. 2019). All three types of models are based on the same probabilistic framework (Blei 2012), but differ in terms of model assumptions. In LDA, topics are assumed to be uncorrelated (Blei et al. 2003). CTM relaxes this assumption and allows topics to be correlated (Blei and Lafferty 2007). STM is similar to CTM in that it allows topics to correlate, but it additionally introduces the use of variables derived from document metadata in the estimation of the topic model (Roberts et al. 2019). This extension allows researchers to investigate the relationship between topics and document metadata.

Document variables derived from metadata typically represent observed characteristics of a specific document (e.g., the time, geographical location, and characteristics of the respondents). A document variable can be implemented in three ways: as a topical prevalence covariate, as a topical content covariate, or both (Roberts et al. 2019). Topical prevalence refers to how much of a document is associated with a topic (Roberts et al. 2019). Including such a covariate in the model allows the frequency of a topic to vary with the prevalence covariate. Topical content refers to the words used within a topic (Roberts et al. 2019). Including such a covariate in the model allows the word rate usage within a topic to vary with the content covariate. This extension therefore allows researchers to investigate meaningful associations between topics and covariates, a strength of STM that distinguishes it from other topic models. For a detailed breakdown of the implementation of STM, I refer readers to the articles published by the developers of the STM framework (Roberts et al. 2014, 2016, 2019).

The opportunity to investigate the relationship between topics and employer size makes STM optimal to explore whether topical content and prevalence can reveal variable associations that better explain the inconsistent findings regarding the relationship between employer size and labor market outcomes for vulnerable groups. Additionally, STM has been used in multiple fields of research, including political science (Roberts et al. 2014), health and medicine (Wright et al. 2022), education (Littenberg-Tobias et al. 2021), and information management (Sharma et al. 2021), demonstrating its wide applicability.

4.1 Preprocessing

Prior to the analysis, all documents were preprocessed. All questions asked by the interviewers were removed from the transcripts to ensure that only the views of the interviewees were expressed in the documents. The interviews were then translated into English using Google Translate. All original interviews and translated interviews were cross-checked to ensure that terms were translated equivalently across documents. The documents were then compiled into a corpus (a collection of written texts) in R. To facilitate the investigation of topical content and prevalence by organization and size, all interviews conducted within the same organization were grouped to set the unit of analysis at the organizational level.

For preprocessing, I used Quanteda, which can be used universally for different topic models (Benoit et al. 2018). First, all punctuation and special characters were removed, and words were converted to lowercase. Next, very common words, called stop words, were removed. These are words that do not contribute to the overall understanding of a document, such as *we*, *me*, *this*, and *that*. I used two different stop-word libraries: one called "Snowball" (Porter 2001) and one called "Smart", both of which are available in Quanteda (Benoit et al. 2018) Additionally, I compiled a custom stop-word library containing names of people, places, and businesses; descriptive words used by the transcribers; and any other common and non-descriptive words not covered by the former stop-word libraries, for example *put*, *either*, and *yet*. The words for the custom library were identified based on word frequency. Additionally, all words were stemmed, or reduced to their roots. The Quanteda package uses Porter's stemming algorithm and the C libstemmer library, generated by Snowball (Benoit et al. 2018; Porter 1980, 2001). The stemming algorithm groups all words that originate from the same stem under the same term, which increases the document's cohesiveness and limits the chance of different versions of the same word appearing multiple times within the same topic. For example, *work*, *working*, *works*, and *worked* would all be gathered under "work." Finally, the lower and upperr limits for word frequency were set to 0.01 and 0.95 across all documents. These limits help prevent extremely rare and very common words from becoming predictive of topics; if the same word is repeated across topics, it becomes more difficult to interpret and differentiate them. The preprocessing yielded 3,562 unique words.

4.2 Selection of number of topics

Before model estimation, the topic modeling algorithm requires the number of topics to be modeled, represented by k, to be input. Selecting an optimal value for k is important, as the number of topics modeled will affect the analysis and results (Sbalchiero and Eder 2020). Setting k too low can produce topics that are too broad, while setting k too high can result in many similar, narrow topics (Greene et al. 2014). There are multiple ways of finding the optimal k for a topic model. A few methods that automatically estimate optimal k for a set interval of topics are available through the STM package in R (SelectModel and SearchK; (Roberts et al. 2019). I adapted a method for estimating multiple topic models with k ranging between three and fifteen, and evaluated the resulting models against five criteria for model fit: held-out likelihood, residual dispersion, lower bound, semantic coherence, and exclusivity (Silge 2018). The first three are related to statistical model fit, while the last two are related to producing topics that are understandable to humans.

Held-out likelihood evaluates the predictive performance of topic models. Random parts of the documents are excluded from the model estimation, and the topic model is trained on the remaining parts. The excluded parts are then used to evaluate the model's predictive power (Roberts et al. 2019; Wallach et al. 2009). A better-fitting model will attribute a higher probability to the excluded part, indicating that the predictive power of the model is high (Wallach et al. 2009). Residual dispersion can be used as an indication of whether too few topics have been specified in the topic model. When the topic model is correctly specified, the residual dispersion should be equal to one (Taddy 2012). This criterion is generally hard to satisfy, so aiming to minimize residual dispersion while satisfying other model-fit criteria can be useful (see the STM documentation). The lower bound is a measure of the convergence of the model. The model is considered to have converged when the bound exhibits sufficiently small changes between iterations.

Semantic coherence is related to the human understanding of topics. It is high when the semantic similarity between high-scoring words within a topic is high (Mimno et al. 2011). For example, a topic concerning children's education that consisted of the words *school*, *child*, *homework*, *teacher*, and *learn* would have relatively high semantic coherence. This measurement helps distinguish topics that result from statistical inferences from those that are semantically understandable. The correlation between the interpretability of topics for humans and the semantic coherence metric is high (Mimno et al. 2011). Exclusivity reflects whether a high-scoring word within one topic also appears as a high-scoring word in other comparable topics (Airoldi and Bischof 2016; Bischof and Airoldi 2012). Topics that score high on exclusivity can be easier to interpret in terms of topical content because they consist of unique words. Given the previous example, the topic would be harder to interpret if *homework*, *teacher*, and *school* were the top predicted words in another topic as well. To achieve high topic quality, both high semantic coherence and high exclusivity are desirable. The two metrics are anti-correlated, however, and the researcher must often compare multiple models to select the final number of topics for the model.

Figure 1 presents the held-out likelihood, lower bound, residual dispersion, and semantic coherence of STM for values of k ranging from 3 to 15. The held-out likelihood is -7.5 at $k=3$, after which it decreases steadily until $k=11$. At $k=12$, the held-out likelihood drops to its low point of -9.5, where it remains up to $k=15$. The residual dispersion never reaches one, which means that the model is not perfectly fitted. The residual dispersion reaches a high point of 38 for $k=6$ but varies between 0 and 5 for the remaining values of k. Semantic coherence is highest at $k=3$, after which it decreases in uneven fluctuations until $k=11$. After $k=11$, semantic coherence decreases steeply and remains between -14 and -16. The lower bound indicates that all topic models converged. Based on this information, the optimal value of k is likely between four to eight topics. For this range, the held-out likelihood is high, the semantic coherence varies between -8 and -9, and the residual dispersion is relatively consistent, except when $k=6$.

Next, I compared the overall mean semantic coherence and exclusivity of the topics within each model with four to eight topics. Figure 2 presents the results. The top-right corner represents the most desirable outcome, in which both semantic coherence and exclusivity are high. The bottom-left corner represents the least desirable outcome, in which both semantic coherence and exclusivity are low. The variation among models in terms of mean semantic coherence and exclusivity is low. All models have semantic coherence values between -9.4 and -8.9 and exclusivity values between 8 and 9. Based on the figure,
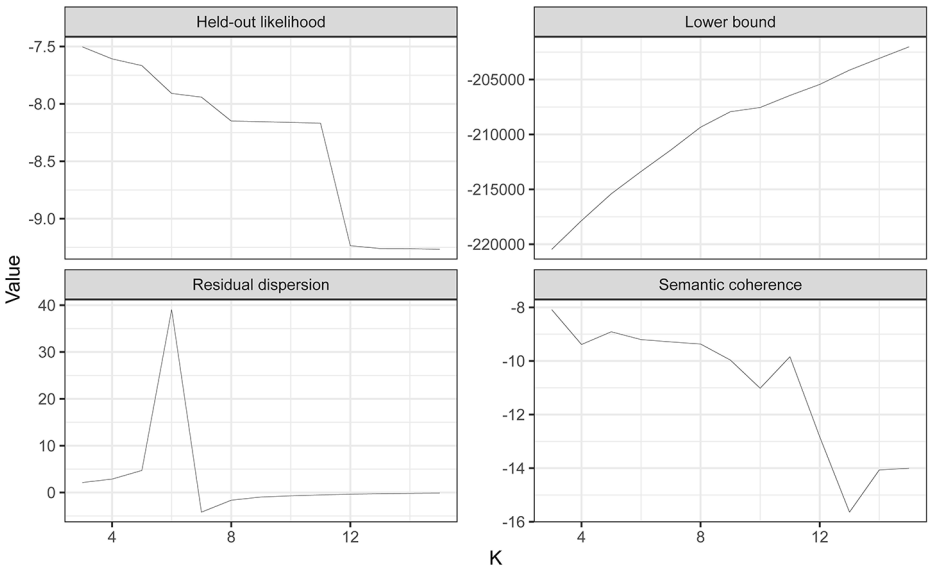
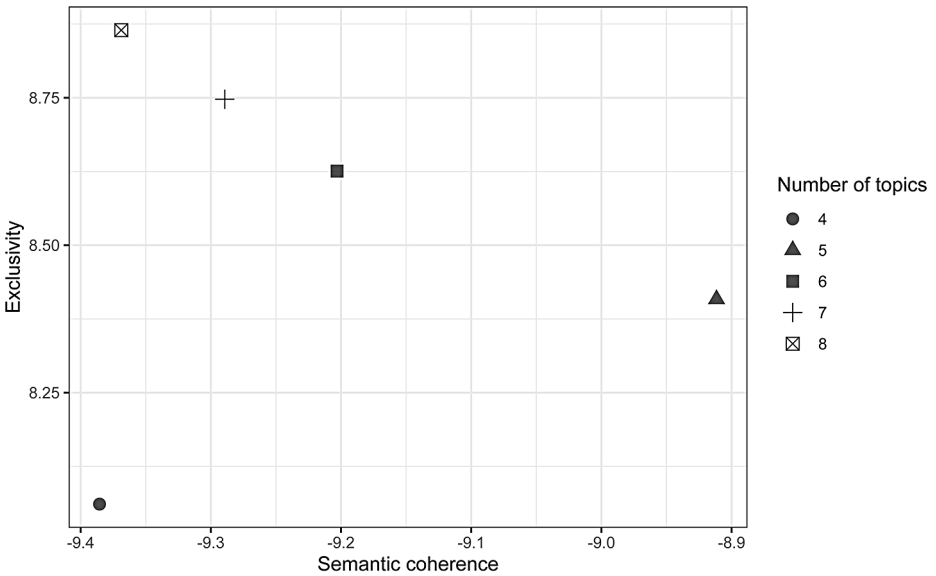**Fig. 1** Selection of the number of topics (k)

Fig. 2 Comparing the mean semantic coherence and exclusivity of different topic models

topic models with 5, 6, 7, or 8 topics perform best. The topic model in which $k=6$ was excluded because of extremely high residual dispersion (see Fig. 1). The topic model in which $k=8$ has the highest mean exclusivity but lower mean semantic coherence than the models in which $k=5$ and $k=7$. The topic model in which $k=7$ has slightly higher mean exclusivity than the model in which $k=5$ but lower mean semantic coherence. Based on this

data, I selected for analysis the topic model in which $k=5$. For an overview of the quality of individual topics within the topic model in which $k=5$, please see S1 in the supplementary material.

4.3 Model specification and analysis

Document variables and initialization types can be specified when estimating a structural topic model. I included employer size—defined as micro (<10), small (≥10 and <50), medium (≥50 and <250), or large (≥250)—as a topical prevalence covariate in the model estimation. In effect, the inclusion of this covariate allows the frequency of topics within the interview transcripts to vary with employer size. Additionally, the results produced by the estimation procedure can be sensitive to the initialization type, which is related to the starting values of the parameters to be modeled, such as the distribution of words linked to a particular topic (Roberts et al. 2019). To ensure stable results, I used spectral initialization, which will produce the same model results regardless of the seed set for the model estimation (Roberts et al. 2019).

After specifying the model, I analyzed the top words associated with each of the five topics to identify their topical content. The STM package has a built-in function that retrieves the documents most representative of a specified topic (findThoughts; (Roberts et al. 2019). I used this function to display the interview transcripts with the highest topic proportions for each topic, thereby determining the organizations that were most representative of each topic. The interview transcripts were then cross-referenced with the top predicted words for each topic and the contexts in which they were used. The analysis yielded topic labels, topic descriptions, and quotes that represented the content of each topic. Finally, topical prevalence by employer size was extracted and visualized (Roberts et al. 2019). To this end, regression analyses were simulated in which the documents represented the units, employer size was the covariate, and the outcomes were the expected topic proportions (see estimateEffect in the STM documentation for details). The expected mean proportion of each topic by employer size was then visualized based on the regression estimates (see plotEstimateEffect in the STM documentation for details).

4.4 Validation

As a form of validation, I estimated topic models in which $k=6$ and $k=7$ and cross-referenced the topical content and order of top predicted words with the model in which $k=5$. For an overview of the top ten predicted words in the topic models in which $k=6$ and $k=7$, please see S2 in the supplementary material. For the model in which $k=6$, Topics 1 and 5 were exact matches for the same topics in the model in which $k=5$, both in terms of topical content and the order of top predicted words. Topics 2, 3, and 4 were partial matches for the equivalent topics in the model in which $k=5$, containing between seven and nine similar words. Topic 6 was a partial match for Topics 3 and 4 in the model in which $k=5$, with three and four similar words, respectively. For the model in which $k=7$, Topic 1 was an exact match with that of the model in which $k=5$, both in order of top predicted words and topical content. Topics 2 and 3 were partial matches for the same topics in the model in which $k=5$, with nine similar words. Topics 4 and 5 had partial matches with the same topics in the model in which $k=5$, with seven and four similar words, respectively. Topic 6 had partial

matches with Topics 3 and 4 in the model in which $k=5$, with three and four similar words, respectively. Topic 7 was a partial match with Topic 5 in the model in which $k=5$, with eight similar words. Overall, these results indicate that the topical content was relatively stable across topic models.

5 Results

5.1 Topical content

This section presents the results of the structural topic model with five topics. Each topic was labeled according to its core content. Table 2 presents the top ten words for each topic in descending order of predicted probability.

5.1.1 Topic 1 – training

Topic 1 was one of the least prevalent topics in the interviews, with topical prevalence of 11.73%. Topic 1 consisted of words that describe training and follow-up in the workplace, including *school, sell, give, customer, apprentice, task, simple, fun, and boss*. Another of the top words, *sudden*, was more difficult to interpret in light of this topic. The most representative organizations for this topic were the gadget store (large) and machinery manufacturer (micro). In these organizations, the daily managers discussed how they try to show their employees that working can be fun; they give the employees suitable training in performing work tasks and offer individual support and follow-up when needed. With the masonry company, candidates for work training were working a few days a week while taking language and professional courses at school. At the machinery manufacturer, tasks were adapted and simplified to fit the skills and needs of employees from vulnerable groups. At the gadget store, the daily manager taught sales skills by instructing one of the included employees in conversation:

> We kind of had to work a lot with conversations and such. But if I managed to explain something to him, why I wanted him to do something a certain way with good arguments, he was always like, "Okay then, I'll try it your way". And then, finally, he started to sell as well. (Daily manager, gadget store)

Table 2 Topical prevalence and content by top predicted words

Topic	Prev.	Top words
1 – Training	11.73	School, sell, give, sudden, customer, apprentice, task, simple, fun, boss
2 – Practicalities	17.65	Department, product, include, place, certificate, apprentice, bring, project, motivate, problem
3 – Recruitment	23.50	Candidate, interview, give, wage, happen, subsidy, colleague, problem, close, participate
4 – Contexts	23.53	Place, leader, open, young, municipality, department, practice, home, give, education
5 – Work demands	23.56	Customer, practice, boss, candidate, language, Norwegian, include, task, department, school

*Prev. = Prevalence. The top words are listed in descending order of predicted probability

5.1.2 Topic 2 – practicalities of the inclusion process

Topic 2 had a topical prevalence of 17.64%. The core content of this topic was the practicalities of the inclusion processes, and it included the words *department, include, place, certificate, apprentice, bring, project, motivate*, and *problem*, all of which relate to the process of including an employee from a vulnerable group in the organization. In this topic, *product* was more difficult to interpret. The most representative organizations for this topic were the bakery (medium), technology company (large), and masonry company (medium). The managers from these organizations discussed how they find places for employees from vulnerable groups who do not fit in elsewhere, how many employees from vulnerable groups they had, how the employees could get apprentice certificates, and what projects they were involved in with NAV (public employment and social services). They also noted the importance of motivating colleagues and employees from vulnerable groups to participate in the inclusion process. Such processes require managers and staff to invest their time, as the production manager of the bakery described:

> Instead of making money from it, we kind of spend money on me spending time for several months on the people we get in here, because nothing works on its own. (Production manager, bakery)

5.1.3 Topic 3 – the recruitment process

Topic 3 was one of the most prevalent, with a topical prevalence of 23.50%. Topic 3 consisted of words that describe the recruitment process: *candidate, interview, give, wage, subsidy, colleague, problem*, and *participate*. The words *close* and *happen* were more difficult to interpret in this topic. The most representative organizations for this topic were the fireplace installer (micro), gardener (micro), landscaper (small), and café (micro). These employers and employees discussed the nature of the hiring process and support from NAV, from the initial introduction to the company by counselors, job coaches, or advisors from NAV to the interview and the follow-up during work training or trial periods. As an employee from a vulnerable group working at the landscaping company described, the interview process differed from previous interview experiences, as the owner tried to get to know the interviewee at both the personal and professional levels:

> Different. The interview itself was okay, but it was really overwhelming since it was very long and it was very elaborate. And he was very accommodating and direct in a way. He did not beat around the bush, as I have experienced many other previous employers doing. (Included employee, landscaper)

5.1.4 Topic 4 – contexts of inclusion

Topic 4 has a topical prevalence of 23.53%. For Topic 4, the core content concerned the contexts of the inclusion processes. The words *place, leader, open, young, municipality, department, practice*, and *give* all describe contexts of inclusion at the individual, organi-

zational, and workplace levels. The inclusion processes described in the interviews were mainly focused on young people, some of these processes were organized through the local municipalities, and the supervisors and workplaces were described as open and social. *Home* and *education* were more difficult to interpret in this topic. The most representative organizations for Topic 4 were the kindergarten (small), IT company (small), logistics company (medium), and nursing home (medium). The HR manager of the IT company noted that the people they included were mostly young men who had dropped out of school because of gaming:

> We have those from [private employment agency]. They are usually young boys who have ended up outside [the labor market] because, for example, they have played [videogames] too much. (HR manager, IT company)

5.1.5 Topic 5 – work demands

Topic 5 was the most prevalent topic in the interviews, with topical prevalence of 23.56%. In Topic 5, the core content was work demands. The words *customer, practice, boss, candidate, language, Norwegian, task*, *school*, and *department* all describe the demands or requirements at work. *Include* was more difficult to interpret in this topic. The most representative organizations for this topic were the three supermarkets (small) and the facility services (large). Managers and employees from vulnerable groups discussed how developing language skills, practicing tasks, and communicating with customers are key to success at work. Several supermarket employees belonging to vulnerable groups attended Norwegian courses a few days each week while receiving work training. The daily manager at Supermarket 2 paid for a Norwegian course for one employee from a vulnerable group, describing it as a "win–win" situation. As explained by an employee from a vulnerable group who worked at Supermarket 1, the boss had work demands linked to the Norwegian language and familiarity with the store and products:

> For example, for me, who is not a Norwegian, he [the daily manager] says that I have to work more and more with the language and develop. More information about the store, about goods, about the rules and such. He also taught me more. (Included employee, Supermarket 1)

5.2 Topical prevalence by employer size

Figure 3 presents the mean topical prevalence by employer size. There are clear differences in mean topic proportions between employers of different sizes. Topic 1, which concerned training, was most prevalent among small and medium-sized employers, with mean topic proportions of about 23% and 35%, respectively. For Topic 2, which had to do with the practicalities of the inclusion processes, the mean topic proportions were about 50% for micro employers, 30% for medium employers, and 3% for small employers. Topic 3, which concerned recruitment, was most prevalent among small employers, with a mean proportion of about 75%, and somewhat less prevalent among large employers, at 18%. For Topic 4,
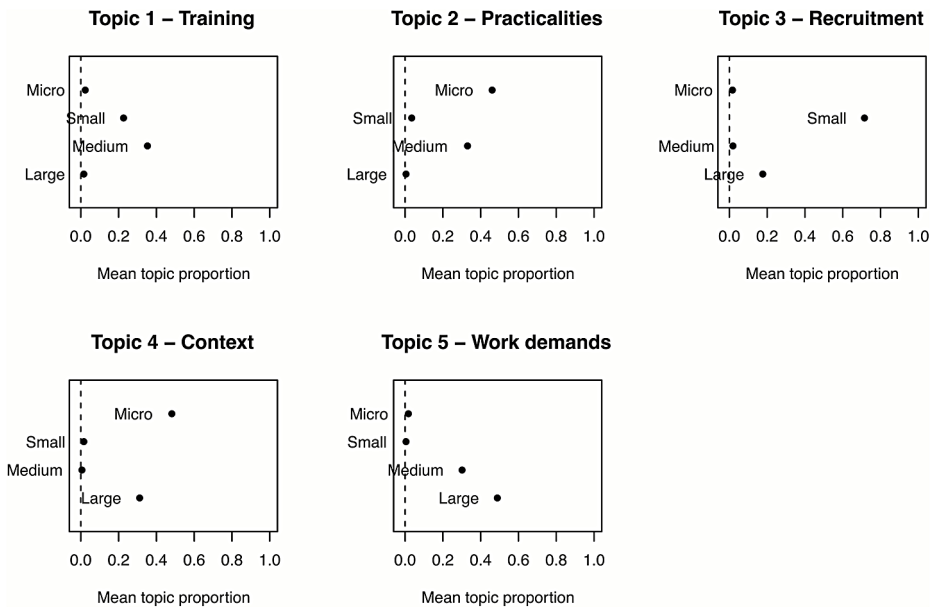
Fig. 3 Mean topic proportions by employer size

which concerned the contexts of inclusion, the mean proportion was around 50% for micro employers and 30% for large employers. Topic 5, focusing on work demands, was most prevalent among large employers, with a mean topic proportion of about 50%. For medium employers, the mean topic proportion was about 37%.

6 Discussion

This study investigated what topics were addressed by employers, whether topical prevalence differed according to employer size, and illustrated how and whether STM could be used as an MMR design in the analysis of qualitative interview data. First, the model estimated five topics based on interviews about sustainable employment for people from vulnerable groups. These were training, practicalities of the inclusion process, recruitment, contexts of inclusion, and work demands. The content of these topics represents issues central to successful labor market integration and sustainable employment for people from vulnerable groups. Similar findings have been reported in previous research, which has identified that fit between person and work tasks, practices for recruitment and selection, contexts of inclusion at different levels, and development and training can affect labor market outcomes for vulnerable groups (Beatty et al. 2019; Hulsegge et al. 2022; Jansen et al. 2021; Kersten et al. 2023; Strindlund et al. 2019). These topics can inform practitioners in placing and matching people from vulnerable groups with prospective employers and facilitate successful work integration. The three most prevalent topics, Topics 3, 4, and 5, were almost equally prevalent across documents, meaning that recruitment, contexts of inclusion, and work demands were the issues most frequently discussed by employers and employees from vulnerable groups. Topic 1, concerning training, was the least prevalent across docu-

ments. Training has been recognized as one of the most important factors underlying success in hiring and integrating employees from vulnerable groups (Kersten et al. 2023). The low prevalence does not necessarily mean that training is less important, but rather reflects that the sample of organizations in this study placed more weight on the other topics.

There were clear differences in the expected topical prevalence by employer size. This can indicate that the topics represent differences in employer knowledge, capability, and motivation to hire and retain people from vulnerable groups (Nagtegaal et al. 2023). For micro employers, the topics concerning contexts of inclusion and practicalities of the inclusion processes were most prevalent. For small employers, the topics of training and recruitment were most prevalent. For medium employers, the topics regarding training, practicalities of the inclusion processes, and work demands were most prevalent. For large employers, the topics concerning recruitment, contexts of inclusion, and work demands were most prevalent. The differences in topic prevalence suggest that organizations of different sizes have distinct practices related to hiring and sustaining people from vulnerable groups. Moreover, the variation in topical prevalence and content may indicate associations between variables that help clarify the inconsistency of previous findings regarding the relationship between employer size and labor market outcomes for vulnerable groups. Operationalizing and quantifying the topics as variables allows them to be used to explain the effect of employer size on labor market outcomes for vulnerable groups. The most prevalent topics for each employer size can likely explain at least part of the association between size and labor market outcomes for vulnerable groups, as these topics represent what is considered to be important to employers in hiring and management of vulnerable groups.

Second, the study introduced topic modeling as an MMR design for integrating qualitative and quantitative methods in the analysis of interview data. Topic modeling has typically been used for shorter textual sources (e.g., social media posts, newspaper articles, abstracts, or open-ended survey responses) and large sample sizes (Macanovic 2022; Roberts et al. 2014; Rohrer et al. 2017). What has been unclear from the literature is whether these methods can be used for smaller samples and on longer texts, both of which are typical of interview data. The successful identification of the content of five topics demonstrates that topic modeling can be used to summarize and identify the core content of interview transcripts. In addition, the possibility of integrating covariates derived from document metadata into model estimation offers great potential for future MMR (Roberts et al. 2019). The quantitization of qualitative data has previously involved rather simple statistical analyses in which dichotomized or categorical variables derived from qualitative data are used in combination with quantitative datasets (Banha et al. 2022; Cabrera and Reiner 2018; Cox et al. 2021; Wao et al. 2011). The integration of covariates into model estimation allows the researcher to directly estimate associations between topics and covariates, making the use of quantitative datasets unnecessary. Any given participant characteristic can be modeled as either a topical prevalence or content covariate for which associations can be extracted and investigated (Roberts et al. 2019). STM thus constitutes an MMR approach in which statistical analysis can be conducted without sacrificing the richness of the data (Driscoll et al. 2007). This can facilitate more in-depth analyses of previously unanalyzed or under-analyzed data, potentially revealing patterns that might otherwise have remained undiscovered (Driscoll et al. 2007).

6.1 Limitations

Although STM has great potential as an MMR design, the method has several potential limitations. First, quantitization of qualitative data material has been criticized for possibly misrepresenting results, as qualitative data samples are often much smaller than quantitative samples (Love and Corr 2022; Maxwell 2010; Pratt 2009). As such, researchers should be cautious about generalizing the results from STM when small samples of qualitative data are used. For small samples, model estimation could be more sensitive to variations in document length as in such cases, longer documents could become predictive of the estimated topics. With larger samples, however, this would likely be less of a problem, as less weight would be ascribed to each document. Alternatively, textual data sources of similar length can be used to limit potential estimation bias arising from differences in document length (e.g., journal abstracts and social media posts with word limits).

Second, quantitization of qualitative data has been criticized for potentially underrepresenting the richness of the data, which is also a possible pitfall for STM (Love and Corr 2022; Maxwell 2010; Pratt 2009). Specifically, in STM, the contexts in which words are used are lost when unigram models are specified. Interpretations of single words drawn from interview transcripts are not based upon the contexts in which the interviewees used the words. The contexts of the words can only be determined by studying the interview transcripts. Researchers should therefore exercise caution in interpreting and determining topical content based on the top predicted words without integrating the interpretations with the full documents. For relatively small sample sizes, researchers can cross-reference top predicted words with the full documents to help interpret the topics. For larger sample sizes, the findThoughts function in the STM package can be used to identify the documents most representative of specific topics, which can help in determining the core content of the topics (Roberts et al. 2019). Alternatively, for research in which the context of a word's use is important, both descriptive textual analysis and topic modeling that considers n words (n-grams) in addition to the focal word can be conducted (Wang et al. 2007; Welbers et al. 2017).

Finally, even when the selection of k is based on criteria that emphasize both statistical fit and human interpretability, the topics can contain words that are difficult to interpret in the estimated topics. This presents a challenge for the use of topic modeling in general, and as an MMR design specifically. Many researchers have suggestions for how to the find optimal number of k topics based on the sample size, length of text chunks, or estimation algorithms (Gan and Qi 2021; Greene et al. 2014; Sbalchiero and Eder 2020; Vayansky and Kumar 2020). As the many suggestions can be hard to navigate, some of the topic modeling packages in R offer automated functions for finding optimal k. These include SelectModel and SearchK for STM (Roberts et al. 2019) and FindTopicsNumber for LDA (Grün and Hornik 2011). Ultimately, it will be up to the individual researcher to select a method for optimizing k that fits both the data and research question.

7 Conclusion

This study contributes to the literature on sustainable employment for vulnerable groups and MMR. Using STM on interview transcripts resulted in the identification of five topics representing key factors that can contribute sustainable employment: training, practicalities of the inclusion process, recruitment, contexts of inclusion, and work demands. The variation in topical prevalence between employer sizes suggests that organizations of different sizes have distinct practices for hiring and integrating people from vulnerable groups. This can contribute to explain the inconsistent findings on the association between employer size and labor market outcomes for vulnerable groups. The findings illustrate the effectiveness of using STM in the analysis of interview transcripts, which can facilitate more in-depth analyses of data, potentially uncovering patterns that might otherwise remain undiscovered.

Future research should aim to test whether the topics, used as operationalized and measurable variables, can at least partially explain the association between employer size and labor market outcomes for vulnerable groups. The clear differences in topical prevalence between employers of different sizes suggest that these variables play an important role in determining labor market outcomes for vulnerable groups and could thus contribute to clarifying the inconsistent findings from previous research. Once the effects of these variables are clarified, the results can be used to refocus efforts to develop and implement policies aimed at increasing labor market participation among vulnerable groups, which will contribute to overall sustainable employment.

The use of STM on qualitative data has great potential as an MMR design. This framework provides a method incorporating both analyses based on statistical techniques (e.g., descriptive statistics and content distribution) and conventional qualitative analyses (e.g., content and thematic analyses). Additionally, topic modeling of qualitative interviews is a straightforward MMR design that can be applied by both inexperienced and advanced researchers, particularly as there are informative tutorials online. Furthermore, the method offers a wide range of options for structuring and analyzing data and interpreting results (Roberts et al. 2019). To further develop STM as an MMR design, future research should strive to determine the possibilities of analyzing qualitative interviews by applying STM to a variety of textual samples of different sizes, including both unanalyzed and previously analyzed samples. By comparing qualitative analyses made using topic models, researchers can identify the optimal sample sizes—in terms of both textual length and number of texts—for evaluating the accuracy of topic models. Applying topic modeling to previously unanalyzed samples can also increase knowledge on whether topic models based on unfamiliar samples of interview transcripts are sufficient to familiarize researchers with the main content of estimated topics and whether topic quality varies with different sample sizes with regard to human interpretability.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

References

Airoldi, E.M., Bischof, J.M.: Improving and evaluating topic models and other models of text. J. Am. Stat. Assoc. **111**(516), 1381–1403 (2016). https://doi.org/10.1080/01621459.2015.1051182

Bacon, N., Hoque, K.: The treatment of disabled individuals in small, medium-sized, and large firms. Hum. Resour. Manag. (2021). https://doi.org/10.1002/hrm.22084

Banha, F., Flores, A., Coelho, L.S.: Quantitizing qualitative data from Semi-structured interviews: A methodological contribution in the Context of Public Policy decision-making. Mathematics. **10**(19), 3597 (2022). https://doi.org/10.3390/math10193597

Barde, B.V., Bainwad, A.M. An overview of topic modeling methods and tools. International Conference on Intelligent Computing and, Systems, C.: (ICICCS), (2017). (2017)

Beatty, J.E., Baldridge, D.C., Boehm, S.A., Kulkarni, M., Colella, A.J.: On the treatment of persons with disabilities in organizations: A review and research agenda. Hum. Resour. Manag. **58**(2), 119–137 (2019). https://doi.org/10.1002/hrm.21940

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A.: Quanteda: An R package for the quantitative analysis of textual data. J. Open. Source Softw. **3**(30), 774 (2018). https://doi.org/10.21105/joss.00774

Bento, J.P.C., Kuznetsova, Y.: Workplace adaptations promoting the inclusion of persons with disabilities in mainstream employment: A case-study on employers' responses in Norway. Social Inclusion. **6**(2), 34–45 (2018). https://doi.org/10.17645/si.v6i2.1332

Bickman, L., Rog, D.J., Hedrick, T.E.: Integrating Qualitative and Quantitative Approaches to Research. In Handbook of applied social research methods. **2**, 283–317 (2009).

Bischof, J., Airoldi, E.M.: Summarizing topical content with word frequency and exclusivity. Proceedings of the 29th International Conference on Machine Learning (ICML-12), (2012)

Blei, D.M.: Probabilistic topic models. Commun. ACM. **55**(4), 77–84 (2012). https://doi.org/10.1145/2133806.2133826

Blei, D.M., Lafferty, J.D.: A correlated topic model of science. Annals Appl. Stat. **1**(1), 17–35 (2007). https://doi.org/10.1214/07-AOAS114

Blei, D.M., Ng, A.Y., Jordan, M.: I. Latent dirichlet allocation. J. Mach. Learn. Res. **3**(Jan), 993–1022 (2003)

Cabrera, L.Y., Reiner, P.B.: A novel sequential mixed-method technique for contrastive analysis of unscripted qualitative data: Contrastive quantitized content analysis. Sociol. Methods Res. **47**(3), 532–548 (2018). https://doi.org/10.1177/0049124116661575

Cooper, A.F., Hankins, M., Rixon, L., Eaton, E., Grunfeld, E.A.: Distinct work-related, clinical and psychological factors predict return to work following treatment in four different cancer types. Psycho-Oncology. **22**(3), 659–667 (2013). https://doi.org/10.1002/pon.3049

Cox, K., Lambert, R., Hitchcock, J.H.: Multiple Linear regression analysis with qualitative data that have been Quantitized. In: The Routledge Reviewer's Guide to Mixed Methods Analysis, pp. 77–88. Routledge (2021)

De Witte, H., Pienaar, J., De Cuyper, N.: Review of 30 years of longitudinal studies on the association between job insecurity and health and well-being: Is there causal evidence? Australian Psychol. **51**(1), 18–31 (2016). https://doi.org/10.1111/ap.12176

Driscoll, D.L., Appiah-Yeboah, A., Salib, P., Rupert, D.J.: Merging Qualitative and Quantitative data in Mixed Methods Research: How to and why not. Ecological and Environmental Anthropology (2007)

Faucett, J., Blanc, P.D., Yelin, E.: The impact of carpal tunnel syndrome on work status: Implications of job characteristics for staying on the job. J. Occup. Rehabil. **10**(1), 55–69 (2000). https://doi.org/10.1023/A:1009441828933

Gan, J., Qi, Y.: Selection of the optimal number of topics for LDA Topic model—taking patent policy analysis as an example. Entropy. **23**(10), 1301 (2021). https://doi.org/10.3390/e23101301

Goss, D., Goss, F., Adam-Smith, D.: Disability and employment: A comparative critique of UK legislation. Int. J. Hum. Resource Manage. **11**(4), 807–821 (2000). https://doi.org/10.1080/09585190050075132

Greene, D., O'Callaghan, D., Cunningham, P.: How many topics? stability analysis for topic models. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15–19, 2014. Proceedings, Part I 14, (2014)

Grün, B., Hornik, K.: Topicmodels: An R package for fitting topic models. J. Stat. Softw. **40**, 1–30 (2011). https://doi.org/10.18637/jss.v040.i13

Hannerz, H., Ferm, L., Poulsen, O.M., Pedersen, B.H., Andersen, L.L.: Enterprise size and return to work after stroke. J. Occup. Rehabil. **22**(4), 456–461 (2012). https://doi.org/10.1007/s10926-012-9367-z

Harney, B., Alkhalaf, H.: A quarter-century review of HRM in small and medium-sized enterprises: Capturing what we know, exploring where we need to go. Hum. Resour. Manag. **60**(1), 5–29 (2021). https://doi.org/10.1002/hrm.22010

Høgelund, J., Holm, A.: Worker adaptation and workplace accommodations after the onset of an illness. IZA J. Labor Policy. **3**(1), 1–18 (2014). https://doi.org/10.1186/2193-9004-3-17

Holm, A., Benn, N. V., & Høgelund, J.: Employers' Importance for the Return to Work of Sick-listed Workers. Socialforskningsinstituttet, working paper 06:2007 (2007)

Hulsegge, G., Otten, W., Van de Ven, H.A., Hazelzet, A.M., Blonk, R.W.B.: Employers' attitude, intention, skills and barriers in relation to employment of vulnerable workers. Work. **72**, 1215–1226 (2022). https://doi.org/10.3233/WOR-210898

Hyggen, C., Vedeler, J.S.: Employer Engagement and active labour market policies. Evidence from a Norwegian Multi-method Study. Social Policy Soc. **20**(4), 548–560 (2021). https://doi.org/10.1017/S1474746420000421

Jansen, J., Van Ooijen, R., Koning, P., Boot, C., Brouwer, S.: The role of the employer in supporting work participation of workers with disabilities: A systematic literature review using an interdisciplinary approach. J. Occup. Rehabil. 1–34 (2021). https://doi.org/10.1007/s10926-021-09978-3

Johnston, V., Way, K., Long, M.H., Wyatt, M., Gibson, L., Shaw, W.S.: Supervisor competencies for supporting return to work: A mixed-methods study. J. Occup. Rehabil. **25**(1), 3–17 (2015). https://doi.org/10.1007/s10926-014-9511-z

Kersten, A., Van Woerkom, M., Geuskens, G., Blonk, R.: Organisational policies and practices for the inclusion of vulnerable workers: A scoping review of the Employer's perspective. J. Occup. Rehabil. **33**(2), 245–266 (2023). https://doi.org/10.1007/s10926-022-10067-2

Kherwa, P., Bansal, P.: Topic modeling: A comprehensive review. EAI Endorsed Trans. Scalable Inform. Syst. **7**(24) (2019). https://doi.org/10.4108/eai.13-7-2018.159623

Kocman, A., Fischer, L., Weber, G.: The employers' perspective on barriers and facilitators to employment of people with intellectual disability: A differential mixed-method approach. J. Appl. Res. Intellect. Disabil. **31**(1), 120–131 (2018). https://doi.org/10.1111/jar.12375

Krause, N., Dasinger, L.K., Deegan, L.J., Rudolph, L., Brand, R.J.: Psychosocial job factors and return-to-work after compensated low back injury: A disability phase-specific analysis. Am. J. Ind. Med. **40**(4), 374–392 (2001). https://doi.org/10.1002/ajim.1112

Littenberg-Tobias, J., Borneman, E., Reich, J.: Measuring equity-promoting behaviors in Digital Teaching simulations: A topic modeling Approach. AERA Open. **7**, 23328584211045685 (2021). https://doi.org/10.1177/23328584211045685

Love, H.R., Corr, C.: Integrating without quantitizing: Two examples of deductive analysis strategies within qualitatively driven mixed methods research. J. Mixed Methods Res. **16**(1), 64–87 (2022). https://doi.org/10.1177/1558689821989833

Macanovic, A.: Text mining for social science–the state and the future of computational text analysis in sociology. Soc. Sci. Res. **102784** (2022). https://doi.org/10.1016/j.ssresearch.2022.102784

Marin-Ferrer, M., Vernaccini, L., Poljansek, K.: Index for risk management inform concept and methodology report—version 2017. (2017). https://doi.org/10.2760/094023

Markussen, S., Røed, K., Røgeberg, O.J., Gaure, S.: The anatomy of absenteeism. J. Health. Econ. **30**(2), 277–292 (2011). https://doi.org/10.1016/j.jhealeco.2010.12.003

Maxwell, J.A.: Using numbers in qualitative research. Qualitative Inq. **16**(6), 475–482 (2010)

McCollum, D.: The sustainable employment policy agenda: What role for employers? Local Econ. **27**(5–6), 529–540 (2012). https://doi.org/10.1177/0269094212444571

Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models Proceedings of the Conference on Empirical Methods in Natural Language Processing, Edinburgh, United Kingdom. (2011)

Nagtegaal, R., de Boer, N., Van Berkel, R., Derks, B., Tummers, L.: Why do employers (fail to) hire people with disabilities? A Systematic Review of Capabilities, opportunities and motivations. J. Occup. Rehabil. **33**(2), 329–340 (2023). https://doi.org/10.1007/s10926-022-10076-1

Porter, M.: F. An algorithm for suffix stripping. Program: Electron. Libr. Inform. Syst. **14**(3), 130–137 (1980). https://doi.org/10.1108/eb046814

Porter, M.F.: Snowball: A language for stemming algorithms. Retrieved 17.01 from (2001). http://snowball.tartarus.org/texts/introduction.html

Prang, K.-H., Bohensky, M., Smith, P., Collie, A.: Return to work outcomes for workers with mental health conditions: A retrospective cohort study. Injury. **47**(1), 257–265 (2016). https://doi.org/10.1016/j.injury.2015.09.011

Pratt, M.G.: From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. Acad. Manag. J. **52**(5), 856–862 (2009). https://doi.org/10.5465/amj.2009.44632557

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G.: Structural topic models for open-ended survey responses. Am. J. Polit. Sci. **58**(4), 1064–1082 (2014). https://doi.org/10.1111/ajps.12103

Roberts, M.E., Stewart, B.M., Airoldi, E.M.: A model of text for Experimentation in the Social Sciences. J. Am. Stat. Assoc. **111**(515), 988–1003 (2016). https://doi.org/10.1080/01621459.2016.1141684

Roberts, M.E., Stewart, B.M., Tingley, D., Stm: An R package for structural topic models. J. Stat. Softw. **91**(2), 1–40 (2019). https://doi.org/10.18637/jss.v091.i02

Rohrer, J.M., Brümmer, M., Schmukle, S.C., Goebel, J., Wagner, G.G.: What else are you worried about?– Integrating textual responses into quantitative social science research. PloS One. **12**(7), 1–34 (2017). https://doi.org/10.1371/journal.pone.0182156

Sbalchiero, S., Eder, M.: Topic modeling, long texts and the best number of topics. Some problems and solutions. Qual. Quant. **54**, 1095–1108 (2020). https://doi.org/10.1007/s11135-020-00976-w

Schneider, U., Linder, R., Verheyen, F.: Long-term sick leave and the impact of a graded return-to-work program: Evidence from Germany. Eur. J. Health Econ. **17**(5), 629–643 (2016). https://doi.org/10.1007/s10198-015-0707-8

Sharma, A., Rana, N.P., Nunkoo, R.: Fifty years of information management research: A conceptual structure analysis using structural topic modeling. Int. J. Inf. Manag. **58**, 102316 (2021). https://doi.org/10.1016/j.ijinfomgt.2021.102316

Silge, J.: Training, evaluating, and interpreting topic models. Julia Silge. **0201** (2018). https://juliasilge.com/blog/evaluating-stm/

Stone, D.L., Colella, A.: A model of factors affecting the treatment of disabled individuals in organizations. Acad. Manage. Rev. **21**(2), 352–401 (1996)

Storey, D.J., Saridakis, G., Sen-Gupta, S., Edwards, P.K., Blackburn, R.A.: Linking HR formality with employee job quality: The role of firm and workplace size. Human Resource Management: Published in Cooperation with the School of Business Administration, the University of Michigan and in Alliance with the Society of Human Resources Management, **49**(2), 305–329, (2010). https://doi.org/10.1002/hrm.20347

Strindlund, L., Abrandt-Dahlgren, M., Ståhl, C.: Employers' views on disability, employability, and labor market inclusion: A phenomenographic study. Disabil. Rehabil. **41**(24), 2910–2917 (2019)

Taddy, M., On Estimation, Selection for Topic Models Proceedings of the Fifteenth International Conference on Artificial Intelligence and, Statistics: Proceedings of Machine Learning Research. (2012). https://proceedings.mlr.press/v22/taddy12.html

Törnberg, A., Törnberg, P.: Combining CDA and topic modeling: Analyzing discursive connections between islamophobia and anti-feminism on an online forum. Discourse Soc. **27**(4), 401–422 (2016). https://doi.org/10.1177/0957926516634546

Tsai, C.-J., Sengupta, S., Edwards, P.: When and why is small beautiful? The experience of work in the small firm. Hum. Relat. **60**(12), 1779–1807 (2007). https://doi.org/10.1177/0018726707084914

Tzgakarakis, S.I., Kritas, D.: Mixed research methods in political science and governance: Approaches and applications. Qual. Quant. **57**, 1–15 (2022). https://doi.org/10.1007/s11135-022-01384-y

Ulstein, J.: The impact of employer characteristics on sustaining employment for workers with reduced capacity: Evidence from Norwegian Register Data. Social Policy Soc. 1–16 (2023). https://doi.org/10.1017/S1474746423000027

Van Berkel, R., Ingold, J., McGurk, P., Boselie, P., Bredgaard, T.: Editorial introduction: An introduction to employer engagement in the field of HRM. Blending social policy and HRM research in promoting vulnerable groups' labour market participation. Hum. Resource Manage. J. **27**(4), 503–513 (2017)

Van Dam, K., Van Vuuren, T., Kemps, S.: Sustainable employment: The importance of intrinsically valuable work and an age-supportive climate. Int. J. Hum. Resource Manage. **28**(17), 2449–2472 (2017). https://doi.org/10.1080/09585192.2015.1137607

Van der Aa, P., Van Berkel, R.: Innovating job activation by involving employers. Int. Social Secur. Rev. **67**(2), 11–27 (2014). https://doi.org/10.1111/issr.12036

Van Ooijen, R., Koning, P.W., Boot, C.R., Brouwer, S.: The contribution of employer characteristics to continued employment of employees with residual work capacity: Evidence from register data in the Netherlands. Scand. J. Work. Environ. Health. **47**(6), 435–445 (2021). https://doi.org/10.5271/sjweh.3961

Vayansky, I., Kumar, S.A.: A review of topic modeling methods. Inform. Syst. **94**, 101582 (2020). https://doi.org/10.1016/j.is.2020.101582

Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D.: Evaluation methods for topic models. Proceedings of the 26th annual international conference on machine learning, (2009)

Wang, X., McCallum, A., Wei, X.: Topical n-grams: Phrase and topic discovery, with an application to information retrieval. Seventh IEEE international conference on data miningICDM (2007). (2007)

Wao, H.O., Dedrick, R.F., Ferron, J.M.: Quantitizing text: Using theme frequency and theme intensity to describe factors influencing time-to-doctorate. Qual. Quant. **45**, 923–934 (2011). https://doi.org/10.1007/s11135-010-9404-y

Welbers, K., Van Atteveldt, W., Benoit, K.: Text analysis in R. Communication Methods Measures. **11**(4), 245–265 (2017)

Wright, L., Paul, E., Steptoe, A., Fancourt, D.: Facilitators and barriers to compliance with COVID-19 guidelines: A structural topic modelling analysis of free-text data from 17,500 UK adults. BMC Public. Health. **22**(1), 34 (2022). https://doi.org/10.1186/s12889-021-12372-6