

Scientific summary

March 2022



1. Background

Youth Futures Foundation is an independent, not-for-profit organisation dedicated to improving employment outcomes for young people from marginalised backgrounds. Youth Futures' mission is to narrow employment gaps by investing in evidence generation and identifying effective interventions aimed at improving employment outcomes for young people. For this reason, Youth Futures approached the Center for Evidence-Based Management (CEBMa) to undertake a review of the scientific research literature regarding the impact of practices that employers use to recruit and select young people from marginalised backgrounds. This review presents an overview of the findings.

2. What is a Rapid Evidence Assessment (REA)?

Evidence reviews come in many forms. One of the best-known is the conventional literature review, which provides an overview of the relevant scientific literature published on a topic. However, a conventional literature review's trustworthiness is often low: clear criteria for inclusion are sometimes lacking and studies may be selected based on the researcher's individual preferences. As a result, conventional literature reviews are prone to severe bias. For this reason, 'rapid evidence assessments' (REAs) are used.

An REA is a specific research methodology that aims to identify the most relevant studies on a specific topic as comprehensively as possible, and to select appropriate studies based on explicit criteria. In addition, the methodological quality of the studies included is assessed by independent reviewers using explicit criteria. In contrast to a conventional literature review, an REA is transparent, verifiable and reproducible, and, as a result, the likelihood of bias is considerably smaller.

3. Main question: What does the review answer?

What is known in the scientific literature about the impact of practices that employers use to recruit and select young people from marginalised backgrounds?

This in turn raises further questions:

- 1. What is meant by a marginalised background?
- 2. What practices used to recruit and select young people from marginalised backgrounds are most widely studied/evaluated?

- 3. What is known about the effectiveness and impact of these practices?
- 4. How can bias in the selection of young people from marginalised backgrounds be minimised?

4. Search strategy: How was the evidence sought?

Three databases were used to identify studies. The studies identified were peer-reviewed academic journals published between 1980-2021, with a focus on 2000-2021 for primary studies.

Our search used a combination of terms including, but not limited to, 'recruitment', 'selection', 'youth', 'young', 'marginalised' and 'disadvantaged'.

An overview of all search terms, databases and queries is provided in Appendix I.

5. Selection process: How were publications and papers selected?

Selection of the scientific publications and papers took place in three phases.

First, titles and abstracts of the 889 + 115 scientific publications and the 2,226 papers identified were screened for relevance based on the title and abstract. In case of doubt or lack of information, the publication/paper was included. Duplicate publications/papers were removed. This first phase yielded 194 scientific publications and 60 papers.

Second, the publications/papers were screened for relevance based on the full text. This second phase yielded 42 scientific publications and 21 papers.

Third, the 42 scientific publications were screened using these inclusion criteria:

- 1. Type of studies: Focusing on empirical studies.
- 2. Measurement: Only studies in which the attributes of successful recruitment and selection practices are quantitatively measured.

In addition, the following exclusion criteria were applied:

- 1. Descriptive studies on gaps (wage, employment status, education) between disadvantaged and non-disadvantaged youth.
- 2. Studies on the outcomes (consequences) of youth unemployment.
- 3. Studies on the early antecedents of youth unemployment.
- 4. Studies on outcomes other than employment, such as health status, mental wellbeing, etc.

6. Data extraction: What data was extracted?

From each study, information relevant to the review question, such as year of publication, research design, sample size, population (e.g., industry, type of employees), type of practice, possible moderators or mediators, main findings and effect sizes, were extracted. An overview of all data extracted is provided in Appendix III (scientific publications) and Appendix IV (papers).

This third phase yielded a total number of 27 scientific publications and 8 papers. An overview of the selection process is provided in Appendix II.

7. Critical appraisal: How was the quality of the included studies judged?

Methodological appropriateness

The classification systems of Shadish, Cook and Campbell (2002), and Petticrew and Roberts (2006) were used to determine the methodological appropriateness of the research design of the studies included. Any discrepancies were resolved through discussion or by consulting a third party where necessary. The following levels of appropriateness were used for the classification, where an 'A' indicates a high level of appropriateness, and a 'D' indicates a low level of appropriateness:

			Stu	dy De	sign	
Purpose	Example	RCT	CBA	C / BA	Cross	Qual
Effect, impact	Does A have an effect/impact on B? What are the critical success factors for A? What are the factors that affect B?	A	В	с	D	na
Association	Is A related to B? Does A often occur with B? Do A and B co-vary?	A	A	A	A	na
Frequency	How often does A occur? How many people prefer A?	na	na	na	A	na
Difference	Is there a difference between A and B?	na	na	Α	A	na
Attitude, opinion	What is people's attitude toward A? Are people satisfied with A? Do people agree with A?	na	na	na	A	с
Experience, perceptions, feelings, needs	What are people's experience with A? What are people's feelings about A? What are people's perceptions about A?	na	na	na	В	Α
Exploration, theory building	Why does A occur? Why is A different from B? In what context does A occur?	na	na	na	В	A

RCT = Randomised controlled trial; CBA = Non-randomised controlled before-after study; C = Controlled study; BA = Before-after study; Cross = cross-sectional study; Qual = Qualitative study; na = not appropriate

Methodological quality

To determine methodological quality, all the studies included were systematically assessed based on explicit quality criteria, such as the PRISMA statement (Moher, Liberati, Tetzlaff, & Altman, 2009) and CONSORT statement (Moher, Schulz, & Altman, 2001), the CASP checklists (Critical Appraisal Skills Programme, n.d.), the checklists of the EPPI-Centre (Newman & Elbourne, 2005), and the critical appraisal criteria developed by the Center for Evidence-Based Management. Based on a tally of the number of weaknesses, the trustworthiness was downgraded. The final level was determined as follows: downgrade one level if two weaknesses were identified, downgrade two levels if four weaknesses were identified, etc.

Effect sizes

To determine the magnitude of an effect, Cohen's rule of thumb (Cohen, 1988), was applied. According to Cohen, a 'small' effect is one that is only visible through careful examination. A 'medium' effect, however, is one that is 'visible to the naked eye of the careful observer'. Finally, a 'large' effect is an effect that anyone can easily see because it is substantial.

Outcome of the appraisal: What is the quality of the studies included?

The overall quality of the included scientific publications was moderate to high. Of the 27 empirical studies included, 16 studies were graded level A, indicating a high level of evidence quality. The outcome of the critical appraisal of each study included is reported in Appendix III.

8. Main Findings

Question 1: What is meant by 'disadvantaged' and 'marginalised' youth?

Based on the screening of the included empirical studies and policy papers, we made the following observations:

Observation 1

There is no generally agreed definition of the term 'marginalised' or 'disadvantaged youth'. Many authors of the studies included in this review use these broad terms to refer to many characteristics and needs.

As a result, the studies cover a wide range of populations from a variety of countries that all have their specific characteristics and needs, (for example, young first-time mothers living in a disadvantaged community versus young men with a chronic health conditions).

This means it is impossible to draw general conclusions about what makes an effective recruitment and selection practice. The effect of an approach is dependent on the context, country and needs of the population defined in that particular study.

Observation 2

Notwithstanding the observation made above, two broad categories can be distinguished that should be taken into consideration when recruiting and selecting young people:

Educated and skilled	Young people who are sufficiently educated and who possess good skills but who have a disadvantage on the labour market due to their age, race, colour, gender, sexual orientation, disability, etc.
Undereducated and unskilled	Young people with lower levels of educational attainment and lack (social and/or professional) skills due to a wide range of factors, such as poverty, drug abuse, mental problems, social issues, discrimination, etc.

Question 2: What practices used to recruit and select young people from marginalised backgrounds are most widely studied/evaluated?

Question 3: What is known about the effectiveness and impact of these practices?

Finding 1: The scientific evidence on practices used to *recruit* disadvantaged youth is scarce in both quantity and quality

The outcome of this review indicates that there are only a limited number of scientific studies on practices employers use to recruit

disadvantaged/marginalised youth. However, some of the papers included discuss and/or recommend practices found to be effective in studies in other domains, such as marketing and (health) promotion. An overview of the most relevant findings and recommendations is provided below

Finding 2: Disadvantaged youth face heightened barriers to information access (Level: na)

Finding 3: To obtain employment, disadvantaged youth tend to use informal job search methods, in particular, their own social network (Level: na)

The endeavour to recruit disadvantaged youth starts with job advertisements. In popular business literature, it is often suggested that the best places for employers to post job advertisements are online platforms such as LinkedIn and Indeed, social media, or traditional media such as newspapers. However, some studies (e.g., Buchanan, 2016; Morris, 2015; Janta, 2011) suggest that disadvantaged youth, especially those who are undereducated and unskilled, face heightened barriers to information access due to lack of insight into the job market, literacy issues, lack of self-efficacy, lack of confidence and/or lack of job searching skills. In addition, it was found (e.g., Collins, 2001; Chapple, 2002; Janta, 2011; Klinthall, 2016) that in order to find employment, disadvantaged or marginalised groups generally tend to rely on personal contacts and their informal network rather than formal channels. These findings suggest that both traditional and online media may not be the most appropriate information channels to recruit disadvantaged youth, but organisations instead should actively look for places where disadvantaged youth congregate. Some policy papers therefore (e.g., Haque, 2020; Chapple, 2002; Perkic, 2019) recommend the use of local, informal networks or reaching out directly to disadvantaged youth by visiting local communitybased organisations such as youth centres, or simply by talking to disadvantaged youngsters on the streets.

Finding 4: Targeted marketing techniques may be effective tools to enhance the awareness of job opportunities among disadvantaged youth (Level: A / na)

Targeted marketing techniques identify groups who share common needs or characteristics and position services to appeal to and reach these groups. A specific form of targeted marketing is ethnic marketing, which incorporates cultural and ethnic cues such as language, symbols and photographic models from relevant ethnic backgrounds, in the communication strategy (Butt, 2010).

A systematic review based on 36 randomised controlled trials in the domain of social marketing has shown that targeted marketing techniques can be effective tools to reach specific target groups, such as young people, particular ethnic groups or disadvantaged groups (Gordon, 2006). This finding supports the results of studies that demonstrate that targeted marketing interventions in general lead to more positive attitudes as well as greater purchasing intent (e.g., Butt, 2012; Leveton, 1996; Nwanko, 1998; Gordon, 2006). Although this review did not identify quantitative studies in which the effect of targeted marketing techniques on the recruitment of disadvantaged youth was evaluated, several policy papers – including a systematic review – suggest that ethnic-specific approaches, targeting subgroups and using language, graphics and prosocial messages in job advertisements that are appealing to the target group, can be effective tools to reach disadvantaged youth (Collins, 2001; Cooper, 2013; Percic, 2019; Thornley, 2010).

Finding 5: There is strong evidence that selection practices are negatively biased against disadvantaged youth (Level A)

The outcome of this review indicates that there are only a limited number of scientific studies on practices employers use to select disadvantaged/marginalised youth. However, in the past decades a large number of studies consistently demonstrate that selection procedures are often negatively biased regarding ethnicity, colour, gender, age, social background, sexual orientation, disability, etc. (e.g., Dean, 2008; Huffcutt, 1998; Martocchio, 1992; Roth, 2008; Whetzel, 2008; Wozniak, 2011). In addition, many of the included policy papers state that disadvantaged youth face marginalisation in the labour market during the selection process due to (unconscious) bias and negative perceptions from employers (e.g., Cooper, 2013; Haque, 2020; Hasluck, 2007; Morris, 2015). For this reason, the REA was expanded with the following question:

Question 4: How can bias in the selection of young people from marginalised backgrounds be minimalised?

An optimal selection procedure focuses on the candidate's capability to perform the essential tasks and functions of the job as defined in the job description. Unfortunately, interviewer decisions are often affected by many other factors – such as the candidates' skin colour, accent, age or socioeconomic background. This phenomenon is referred to as 'bias': a tendency to have a subjective opinion or view toward or against an

individual, an ethnic group, nation, religion or social class, etc. Biases tend to be unconscious, so employers don't always realise that their selection practices could be biasing the selection outcome (e.g. Kahneman, 1982; Simon, 1965). Numerous high-quality studies have consistently shown, however, that organisational selection practices are vulnerable to subjectivity, biases, prejudice and other influences. This is particularly the case when organisations are committed to hiring, evaluating and promoting young people who could be negatively affected by bias due to their age, race, colour, gender, sexual orientation or disability, rather than youngsters who lack the necessary social and/or professional skills.

In general, there are three ways of minimising bias in the selection process. The first is to promote the job-relatedness of the selection method, that is, to measure skills and abilities needed for the job (also referred to as the validity of the selection method). The second is to promote the consistent use of the method such that the scores a candidate obtains from two administrations of the same selection method are highly correlated (also referred to as the *reliability* of the method). Finally, when selecting employees, subgroup differences must be considered. An overview of the main findings on all three aspects is provided below.

A. Practices that improve the validity of selection methods Finding 6: Structured job interviews are more resistant to bias than unstructured ones (Level C)

Job interviews can be structured, semi-structured or open (unstructured). Structured interviews use a fixed format, where the same questions (prepared beforehand) are asked of each applicant, usually in the same order. More than 13 meta-analyses have consistently found strong evidence that structured job interviews are less biased against marginalised groups than unstructured ones (e.g., Dixon, 2002; Huffcutt, 1999; Huffcutt, 1994; Levashina, 2014; McDaniel, 1994). More specifically, use of structured interviews substantially reduces disparities related to race, gender, age, disability, etc. However, it was found that unstructured interviews become less biased against marginalised groups as the complexity of the job increases, in particular jobs that comprise multiple, partially interdependent subtasks (Huffcutt, 1998).

Finding 7: The use of scoring anchors increases the reliability and validity of job interviews (Level C)

Another factor that is related to the reliability and validity of (structured) job interviews is the use of scoring anchors, also referred to as benchmark answers. In this type of interview, questions and their potential answers are based on a thorough job analysis and then each answer is rated according to its quality/correctness. In the interview session, the interviewers match the interviewee's answers to benchmark answers. Several meta-analyses have shown that the use of scoring anchors increases the validity and reliability of the job interview (Dixon, 2002).

Finding 8: Interviewer training improves an interview's validity (Level C)

Several meta-analyses indicate that the validity of the interview increases if the interviewers have previously been provided with training (e.g., Huffcutt, 1999), for example, in how to conduct a (structured) job interview and how to use scoring anchors. A systematic review found that training did not appear to affect the validity of panel interviews, however, results suggest that it may improve their reliability (Dixon, 2002).

Finding 9: Using the same interviewer(s) across all applicants improves an interview's validity (Level D)

A meta-analysis based on 120 studies demonstrated that an interview's validity moderately increases if all applicants are interviewed by the same person or panel (Huffcutt, 1999). It was found that this practice is especially important when the interview is not highly structured.

Finding 10: The use of situational judgment tests improves the validity of job interviews (Level A)

Situational judgment tests (SJTs) present work-related situations and then ask the applicant about possible responses to that situation. There are typically three types of SJT questions: situational, past behaviour and knowledge. Situational questions ask respondents how they would likely behave in a given situation (e.g., ''Assume that you were faced with the following situation ... what would you do?''). Past behaviour questions ask respondents about how they behaved in a past situation (e.g., ''Can you think of a time when ... what did you do?''). Finally, knowledge questions ask respondents to evaluate the effectiveness of possible responses to a given situation. A large number of studies have indicated that the use of SJTs substantially improves the validity of job interviews (e.g., McDaniel, 2007). A meta-analysis found that past behaviour questions in combination with anchored answering scales (above) tend to yield a higher interview validity than situational questions, especially when used for low complexity jobs (Taylor, 2002).

B. Predictors of job performance

Finding 11: General Mental Ability in combination with a structured interview or a work sample test is the most valid and reliable predictor of future performance (Level C)

Finding 12: Age and job experience are unreliable predictors of future performance (Level C)

A meta-analysis based on 85 years of research in personnel selection suggests that the validity of General Mental Ability (GMA), also referred to as General Intelligence (IQ), measures for predicting job performance is stronger than that of any other method. Further, when used together with a work sample test or structured interview, GMA may be an even better forecaster of future performance. On the other hand, age and job experience, emphasised in many selection processes, are <u>not</u> good performance predictors (Schmidt, 1998). This finding was confirmed by an unpublished meta-analysis that included 100 years of research (Schmidt, 2016).

Finding 13: In general, assessments centres are unreliable predictors of future performance (Level A)

An assessment centre is a recruitment selection process where an organisation typically assesses a group of candidates at the same time and place using a range of selection exercises.

Many organisations use assessment centres because they believe assessment centres provide them with a wide range of insights about candidates and their developmental possibilities. However, many studies have demonstrated that assessment centres have little incremental validity over GMA measures (a 2% increase), work sample tests and structured interviews (Schmidt, 2016, see also Lance, 2008).

C. Subgroup differences

The above findings constitute the best available evidence on the topic of employee selection. Given the target population of this REA, that is

disadvantaged and marginalised youth, the studies included were scanned for ethnic, gender, age and other relevant subgroup differences.

The following findings draw on studies from a range of different countries. They provide a wider global context. Caution should be applied in transferring findings across national settings into the UK due to different recruitment practices and population demographics.

Finding 14: Well-designed general cognitive ability tests account, and adjust, for possible subgroup differences (Level A)

It is widely agreed that cognitive ability tests are the most valid and reliable predictor for future performance (see above). However, evidence shows that these tests can have serious flaws. Numerous studies indicate that White applicants obtain higher scores than ethnic minority applicants, suggesting a test bias against minority groups (e.g., Martocchio, 1992; Schmidt, 1988; Roth, 2001). For example, early meta-analyses on this topic found that White applicants tend to score about one standard deviation higher than Black applicants and about two-thirds of a standard deviation higher than Hispanic applicants (e.g., Gottfredson, 1988; Huffcutt, 1998; Hunter, 1984; Sackett, 1994). Obviously, such differences can have profound effects on hiring outcomes. Well-designed cognitive ability tests therefore account (and adjust) for possible group differences.

Finding 15: Structured interviews have lower ethnic group differences than non-structured interviews (Level A)

It was found that, on average, high-structured interviews result in lower ethnic group differences than low-structured interviews (Huffcutt, 1998). In addition, it was found that these differences tend to decrease as the complexity of the job increases, and when there is a greater proportion of minorities in the applicant pool.

Finding 16: There are differences between minority groups on work sample tests (Level A)

A widely used method for the selection of new employees is the work sample test, that is, a hands-on simulation of the job that must be performed by the applicant. However, when subjective measures of performance are used (rather than objective measures or anchored test scores – see above), White applicants received substantially higher ratings (Martocchio, 1992; Roth, 2008).

Finding 17: There are differences between ethnic groups on situational judgment tests (Level A)

A meta-analysis of 62 studies found that, on average, White applicants perform better on situational judgment tests than Black, Hispanic and Asian applicants (Whetzel, 2008). Female applicants performed slightly better than male applicants.

Finding 18: There are differences between ethnic groups in ratings of assessment centres (Level A)

A meta-analysis of 27 studies found that there are substantial differences between the assessment centre ratings of Black and White applicants. (Dean, 2008). Differences between Hispanic vs White applicants and male vs female groups tend to be lower.

9. Conclusion

This review identified only a limited number of scientific studies on practices employers use to recruit or select disadvantaged youth. Fortunately, the scientific evidence on the recruitment and selection of employees in general is vast and many of the insights that emerge from these studies are also relevant and applicable to the population of disadvantaged youth. When it comes to recruitment practices, the evidence suggests that using local, informal networks, reaching out directly to disadvantaged youth, and using targeted marketing techniques can be an effective recruitment strategy. When it comes to selection practices, the evidence consistently demonstrates that using unbiased, valid and reliable selection methods substantially increases the employment outcomes for young people from marginalised backgrounds. In particular, the use of GMA tests, work sample tests, structured interviews, scoring anchors and situational judgment tests – if they account for group differences - increases the likelihood that the outcome of the selection procedure will be a fair representation of the population from which the organisation recruits its candidates - including young people from marginalised backgrounds. It should be noted that these selection methods are particularly effective when organisations are committed to hiring young people who could be negatively affected by bias due to their age, race, colour, gender, sexual orientation or disability, rather than young people who lack the necessary social and/or professional skills.

10. Limitations

This REA aims to provide a balanced assessment of what is known in the scientific literature about the impact of practices that employers use to recruit and select young people from marginalised backgrounds by using the systematic review method to search and critically appraise empirical studies. However, in order to be 'rapid', concessions were made in relation to the breadth and depth of the search process, such as the exclusion of unpublished studies, the use of a limited number of databases and a focus on empirical research published in the period 2000 to 2019. As a consequence, some relevant studies may have been missed.

A second limitation concerns the critical appraisal of the studies included, which did not incorporate a comprehensive review of the psychometric properties of their tests, scales and questionnaires.

Given these limitations, care must be taken not to present the findings presented in this REA as conclusive.

References

Anderson, N., Salgado, J.F., & Hülsheger, U.R., (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. International Journal of Selection and Assessment, 18(3), 291-304.

Barends, E., Rousseau, D.M. & Briner, R.B. (Eds). (2017). CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations, Version 1.0. Center for Evidence Based Management, Amsterdam. Available from <u>www.cebma.org/guidelines/</u>

Buchanan S & Tuckerman L. (2016). The information behaviours of disadvantaged and disengaged adolescents. *Journal of Documentation*. 72(3):527-48.

Butt, M. M., & de Run, E. C. (2011). Do target and non-target ethnic group adolescents process advertisements differently?. Australasian Marketing Journal (AMJ), 19(2), 77-84.

Butt, M. M., & de Run, E. C. (2012). Can Ethnically Targeted Advertising Work for Malay Adolescents?: The Moderating Role of the Strength of Ethnic Identity. *Asian Academy of Management Journal*, 17(1).

Chapple, K. (2002). "I Name it and I Claim it – in the Name of Jesus, this Job is Mine": Job Search, Networks, and Careers for Low-Income Women. *Economic Development* Quarterly, 16(4), 294-313.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Collins, E. L., & Zoch, L. M. (2001). Targeting the young, the poor, the less educated: thinking beyond traditional media. *Public Relations Review*, 27(2), 197-212.

Cooper Gibson Research & Disability Rights UK. (2013). Supported internship trial for 16 to 24 year old learners with learning difficulties and/or disabilities: an evaluation. Department for Education.

Dean, M.A., Roth, P.L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, 93(3), 685-691.

Dixon, M., Wang, S., Calvin, J., Dineen, B., & Tomlinson, E. (2002). The panel interview: A review of empirical research and guidelines for practice. *Public Personnel Management*, 31(3), 397-428.

Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72(3), 493-511.

Gordon, R., McDermott, L., Stead, M., & Angus, K. (2006). The effectiveness of social marketing interventions for health improvement: what's the evidence? *Public Health*, 120(12).

Gottfredson, L. S. (1988). Reconsidering fairness: A matter of social and ethical priorities. *Journal of Vocational Behavior*, 33(3), 293-319.

Haque,Z., Mompelat,L. & McIntosh,. (2020). Voice of the underrepresented: What's stopping young Black men getting on in the digital and construction sectors? Greater London Authority.

Hasluck, C., Green, A., (2007). What works for whom? A review of evidence and metaanalysis for the Department for Work and Pensions. Social Care Institute of Excellence.

Huffcutt, A.I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79(2), 184-190.

Huffcutt, A.I. & Roth, P.L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology*, 83(2), 179-189.

Huffcutt, A.I., & Woehr, D.J. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior*, 20(4), 549-560.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, 96(1), 72.

Janta H. (2011). Polish migrant workers in the UK hospitality industry. International Journal of Contemporary Hospitality Management. 23(6):803-19.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. Cambridge university press.

Klinthäll, M., & Urban, S. (2016). The strength of ethnic ties: Routes into the labour market in spaces of segregation. *Urban studies*, *5*3(1), 3-16.

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*, 1(1), 84-97.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67(1), 241-293.

Leveton, L., Mrazek, P., & Stoto, M. (1996). Social marketing to adolescents and minority populations. Social Marketing Quarterly, 3(1), 6-23.

Martocchio, J.J., & Whitener, E.M. (1992). Fairness in personnel selection: A meta-analysis and policy implications. *Human Relations*, 45(5), 489-506.

McDaniel, M.A., Hartman, N.S., Whetzel, D.L., & Grubb III, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1).

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L., & Maurer, S.D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599-616.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), e1000097.

Moher, D., Schulz, K. F., Altman, D., Consort Group, & CONSORT Group. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Jama, 285(15), 1987-1991.

Morris, M. (2015). Supporting ethnic minority young people from education into work. Joseph Rowntree Foundation. Retrieved from http://www.scie-socialcareonline.org.uk/supporting-ethnic-minority-young-people-from-education-into-work/r/a11G000000CT0ZbIAL

Nwankwo, S., & Lindridge, A. (1998). Marketing to ethnic minorities in Britain. Journal of Marketing Practice: Applied Marketing Science.

Perkić, D. (2019). Engaging Disadvantaged Youths in Volunteering: What Works in Recruitment? (Master's thesis).

Petticrew, M., & Roberts, H. (2006). How to appraise the studies: an introduction to assessing study quality. Systematic reviews in the social sciences: A practical guide, 125-163.

Roth, P., Bobko, P., McFarland, L., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of Black-White differences in overall and exercise scores. *Personnel Psychology*, 61(3), 637-662.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49(11), 929.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, 33(3), 272-292.

Schmidt, J. A., & Pohler, D. M. (2018). Making stronger causal inferences: Accounting for selection bias in associations between high performance work systems, leadership, and employee and customer satisfaction. *Journal of Applied Psychology*, 103(9), 1001-1018.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2), 262.

Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years. *Fox School of Business Research Paper*, 1-74.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton, Mifflin and Company.

Shaughnessy, J. J., & Zechmeister, E. B. (1985). Research methods in psychology.

Simon, H. A. (1965). Administrative decision making. Public Administration Review, 31-37.

Taylor, P.J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75(3), 277-294.

Thornley, L., & Marsh, K. (2010). What works in social marketing to young people. Systematic Review for the Health Research Council of New Zealand and the Ministry of Youth Development Final Report.

Whetzel, D.L., McDaniel, M.A.; & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3), 291-309.

Wozniak, A. (2011). Field Perspectives on the Causes of Low Employment Among Less Skilled Black Men. The American Journal of Economics and Sociology, 70(3), 811.

Appendix I Search terms and queries

Three databases were used to identify studies: ABI/INFORM Global, Business Source Premier and PsycINFO. The following generic search filters were applied during the search:

- Scholarly journals, peer-reviewed.
- Published in the period 1980 to 2021 for meta-analyses and the period 2000 to 2021 for primary studies.
- Articles in English.

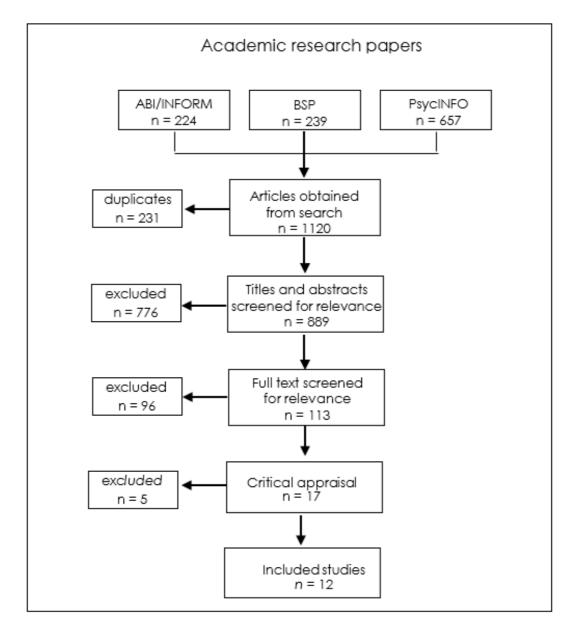
A search was conducted using combinations of various search terms, such as (but not limited to) 'recruitment', 'selection', 'youth', 'young', 'marginalised' and 'disadvantaged'. We conducted seven different search queries and identified a total number of 889 scientific publications.

In addition, an extended search was conducted in PsycINFO for metaanalyses published in the past 40 years on the attributes of effective employee selection. This search yielded another 115 scientific publications.

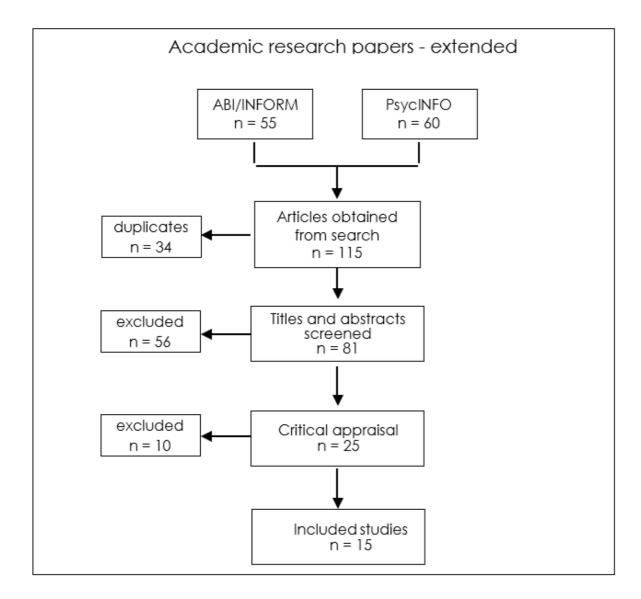
Finally, a search was conducted in Econlit, IBSS, Planex, Social Policy & Practice (Social Policy), Social Science Database and the Sociology Collection to identify unpublished reports, government documents, policy papers and conference proceedings. This search identified a total of 2,226 papers.

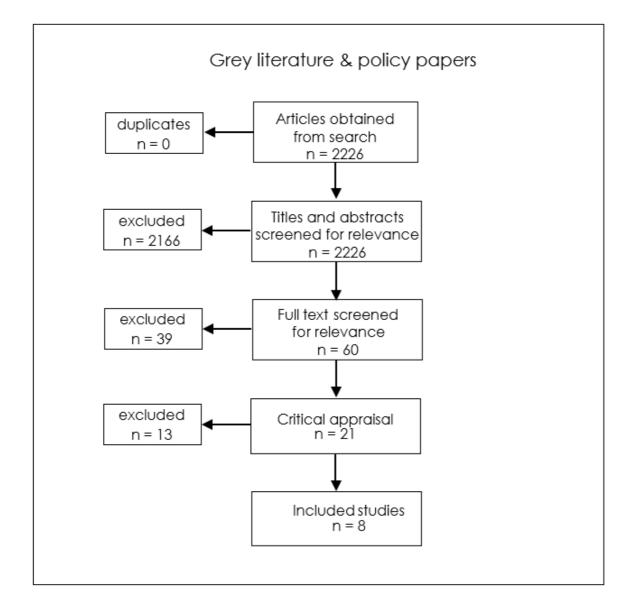
ABI/INFORM Global, Business SourcePremier, PsycINFO peer reviewed, scholarly journals, October 2021								
Search terms	ABI	Р	PSY					
S1: ti(recruit*) OR ti(select*) OR ti(employ*) OR ab(recruit*) OR ab(select*) OR ab(employ*)	389,206	476,906	511,078					
S2: ti("labour market") OR ab("labour market") OR ti("labor market") OR ab("labor market")	23,111	21,137	5,668					
S3: ti(disadvantaged) OR ti(margin*) OR ti(minorit*) OR ti(ethnic*) OR ti(migrant*) OR ti(poor)	19,598	21,993	34,763					

S4: ti(young*) OR ti(youth) OR ti(adolescent*) OR ab(young*) OR ab(youth) OR ab(adolescent*)	49,917	58,360	402,380
S5: S1 AND S3 AND S4, limit > 2000 *filter quantitative studies, age 13 - 29	156	121	506*
S6: S2 AND S3 AND S4, limit > 2000 *filter quantitative studies, age 13 - 29	36	38	16*
S7: S5 OR S6	173	136	513



Appendix II Study selection





Appendix III (scientific publications)

Data extraction form - employee selection (general)

Study level is rated in relation to its intended purpose (see p.5), i.e. to assess: effect or impact (eff); frequency (freq); or difference (diff).

1st Author and year	Design and sample size	Sector / Population	Main findings	Effect sizes	Limitations	Level
1.	Meta-	Job		Not reported	Design of the	А
Anderson	analysis of k	applicants	Quantitative summary into applicant reactions		included studies is	(freq)
(2010)	= 38 studies,	from 17	to popular methods of employee selection.		not reported	
	N = 8974	countries,				
		junior level,	1. Reaction favourability is structurally similar			
		entry-level	across countries and reveals a three-tier			
		job roles and	clustering of overall favourability perceptions –			
		student	most preferred (work samples, interviews), less			
		surrogate	favourably evaluated (resumes, cognitive			
		samples.	tests, references, biodata, personality			
			inventories), and least preferred (honesty tests,			
			personal contacts, graphology).			
			2. Further analyses revealed strong positive			
			correlations between favourability ratings and			
			their validity and international usage.			

2. Dean	Meta-	Job	1. Black–White differences in assessment	1.	Design of the	А
(2008)	analysis of k	applicants or	centres are higher than previously expected.	Black-White group: d =	included studies is	(diff)
	= 27 studies	job		0.52;	not reported	
		incumbents,	2. The adverse impact potential of assessment			
		White, Black,	centres appears to be lower for Hispanic-White	2.		
		Hispanic;	and male-female groups.	Hispanic-White group:		
		male and		d = 0.28; male-female		
		female.		group: d = -0.19.		
3. Dixon	Systematic	Participants		No pooled effect sizes	Design of the	С
(2002)	review	of panel	Results revealed that setting, job analysis,	reported	included studies is	(eff)
		interviews,	scoring anchors, question type, training,		not reported	
		adult	structure, combination method and predictive			
		population	criteria have all demonstrated usefulness in		Synthesis is mostly	
		from wide	explaining variance between (panel) interview		narrative	
		range of	studies.			
		industries.				
4. Gaugler	Meta-	Adult	The validity of assessment centres tend to be		Old study that may	А
(1987)	analysis,	population	higher when	Effect sizes > purpose	require revision.	(eff)
	most likely	from wide		Performance r = .36		
	includes	range of	- the percentage of female assessees is high	Potential r = .53	Design of the	
	experimental	industries.	- assessors were psychologists (rather than	Dimension r = .33	included studies is	
	studies		managers)	Training r = .35	unclear.	
			- peer evaluation was used.	Career r =.36		
	k = 50				Effect sizes	
			Age of assessees and the percentage of	(concerns weighted r's	somewhat unclear	
			minority assessees do not moderate	corrected for statistical		
			assessment centre validities	artifacts)		

			Study design only slightly affected the effect sizes found.			
5. Huffcutt	Meta-	Entry-level	(1) Structure is a major moderator	Not reported	Design of the	С
(1994)	analysis	jobs	of interview validity;		included studies is	(eff)
		applicants.			unclear.	
	K = 114		(2) Interviews, particularly when structured,			
			can reach levels of validity that are		Effect size is not	
			comparable to those of mental ability tests;		reported.	
			(3) Although validity does increase through			
			much of the range of structure, there is a point			
			at which additional structure yields essentially			
			no incremental validity.			
6. Huffcutt	Meta-	Black, White	Interviews as a whole do not appear to affect		Old study that may	А
(1998)	analysis	and Hispanic	minorities nearly as much as mental ability	Black-White group	require revision,	(diff)
		applicants.	tests.	overall d = .25	the result might be	
	K = 31			structure low vs high	not relevant.	
			Results also suggested that (a) high-structure	.32 vs .23		
			interviews have lower group differences on	job complexity low vs	Design of the	
			average than low-structure interviews, (b)	high	included studies is	
			group differences tend to decrease as the	.43 vs 0	unclear.	
			complexity of the job increases, and (c) group			
			differences tend to be higher when there is a	Hispanic-White group		
			greater proportion of a minority in the	overall d = .26		
			applicant pool.	structure low vs high		
				.32 vs .23		

				job complexity low vs		
				high		
				.54 vs23		
7. Huffcutt	Meta-	Adult	The purpose of this study was to analyse the	Corrected correlations	Design of the	С
(1999)	analysis of k	population	relationship between four interviewer-related	1. r = .41	included studies is	(eff)
	= 120 studies,	from wide	factors and the validity of the employment	2. r = .31	unclear.	
	(N = 18,158)	range of	interview.	3. r = .36		
		industries.				
			Results suggested that (1) training should be	4. r =05		
			provided to interviewers regardless of whether			
			the interview itself (i.e., the questions and			
			rating scales) is structured; (2) the same			
			interviewer should be used across all			
			applicants, especially when the interview itself			
			is not highly structured; (3) the interviewer			
			should take notes;			
			(4) using a panel of interviewers does not			
			contribute to the validity, and may actually			
			have a detrimental effect.			
8.	Meta-	US,	The results show that Whites obtained higher	All effect sizes reported	Old study that may	А
Martocchio	analysis,	workplace	scores than non-Whites on cognitive ability	(d's) were medium (in	require revision.	(diff)
(1992)	k = 8	settings,	tests and received higher performance ratings.	the range of .2846)		
		different	However, these differences were greater on		Very small number	
		contexts;	the tests than on the ratings.		of studies.	
		Black, White				
		and Hispanic	In addition, the comparison between			
		participants.	predicted and actual differences on			

supervisory ratings of performance for Whites	De	sign of the
and non-Whites indicate that predicted mean		luded studies is
criterion differences are smaller than the	Und	clear.
actual mean criterion differences, suggesting		
that ratings indicate that Whites outperformed	Sar	nple size (total
non-Whites to a greater extent than was		is unclear.
predicted.		
Also, the samplesized, weighted average		
criterion differences between Whites and non-		
Whites is larger with subjective measures of		
performance than with objective measures of		
performance. The validities between cognitive		
ability and objective and subjective ratings are		
highly similar.		
Note: The actual mean difference between		
Whites and non-Whites on objective criteria		
was only009.		
Thus, the findings imply that test results might		
actually not be fair predictors of performance		
for minorities.		

9.	Meta-	Adult	Interview validity depends on the content of	See table 4, all effect	Design of the	С
McDaniel	analytic	population	the interview (situational, job related, or	sizes reported were	included studies is	(eff)
(1994)	review	from wide	psychological), how the interview is	small to moderate	unclear.	
		range of	conducted (structured vs unstructured; board	(in the range of .24)		
	k = 245	industries.	vs individual), and the nature of the criterion			
	N = 86,311		(job performance, training performance, and			
			tenure; research or administrative ratings).			
			Situational interviews had higher validity than			
			did job-related interviews, which, in turn, had			
			higher validity than did psychologically-			
			based interviews. Structured interviews were			
			found to have higher validity than unstructured			
			interviews.			
			Interviews showed similar validity for job			
			performance and training performance			
			criteria, but validity for the tenure criteria was			
			lower.			
10.	Meta-	Not reported.		SJT - Cogn ability: ρ =	Design of the	А
McDaniel	analysis		Results showed that response instructions	.32	included studies is	(diff)
(2007)			influenced the constructs measured by the	SJT (knowl): ρ = .35	unclear.	
			tests.	SJT (beh): ρ = .19		
			1. Tests with knowledge instructions had higher		Study does not	
			correlations with cognitive ability.	SJT - Big Five: ρ = all in	take into account	
				the range of .2/.3, but	job-performance	
					context (eg. in a	

2. Tests with behavioural tendency instructions	higher for behavioral	high safety/risk	
showed higher correlations with personality	SJT	environment SJT	
constructs.		maybe more	
3. Response instructions had little moderating	Incremental validity for	relevant, whereas	
effect on criterion-related validity.	both cognitive ability	in a situation	
4. Supplemental analyses showed that the	and Big 5 > .1	where the job	
moderating effect of response instructions on		applicant will	
construct validity was not due to systematic		receive training it	
differences in item content.		may be less	
5. SJTs have (very small) incremental validity		relevant).	
over cognitive ability, the Big 5, and over a			
composite of cognitive ability and the Big 5.			
* Situational judgment tests (SJTs) are			
personnel selection instruments that present			
job applicants with work-related situations and			
possible responses to the situations. There are			
typically 2 types of instructions: behavioural			
tendency and knowledge. Behavioural			
tendency instructions ask respondents to			
identify how they would likely behave in a			
given situation. Knowledge instructions ask			
respondents to evaluate the effectiveness of			
possible responses to a given situation.			

11. Nye	Meta-	Not reported.	FROM ABSTRACT	Small	Design of the	С
(2017)	analysis			(.1632)	included studies	(eff)
			Previous studies found that interests predict		not reported.	
	k = 90		performance but the results differed with			
	(1858 ES)		respect to the validity of interest congruence.		Results and	
			This meta-analysis demonstrates that interest		relevance	
			congruence is a stronger predictor of		somewhat	
			performance outcomes than interest scores		unclear.	
			alone, with baseline correlations of 0.32 and			
			0.16, respectively.			
12. Roth	Meta-	Public sector	Data for Black–White ethnic group shows that	Medium to Large	Design of the	А
(2008)	analysis,	organisations;	overall work sample exams differences were		included studies is	(diff)
	K = 40,	Black and	markedly larger for samples of job applicants	d = .36 (incumbents	not reported.	
	N = 8,087	White	(d = .73) than previously thought.	studies)		
		incumbents		d = .73 (early-stage		
		and	There were also substantial differences	applicants studies		
		applicants.	between different exercises, and saturation of			
			different constructs influenced work sample			
			differences. For example, work sample test			
			ratings of cognitive and job knowledge skills			
			were associated with a mean observed d =			
			.80, whereas ratings of various social skills were			
			associated with mean observed ds that varied			
			from .21 to .27.			

13a.	Literature	Not reported.	The two combinations with the highest	1. Multiple R = .63	No serious	С
Schmidt	review of 85		multivariate validity for predicting job	2. Multiple R = .63	limitations.	(eff)
(1998)	years of		performance were 1) GMA plus a work sample			
	research		test and 2) GMA plus a structured interview.	See table 1 for all other		
				(18) predictors		
13b.	Literature	Not reported		Similar as in Schmidt,	Research	С
Schmidt	review of 100		The two combinations with the highest	1998, see table 1 for all	methodology	(eff)
(2016)	years of		multivariate validity for predicting job	other (30) predictors	unclear.	
	research		performance were 1) GMA plus a work sample			
			test and 2) GMA plus a structured interview.		The paper was	
	(replication				never published	
	of Schmidt,				(reason?).	
	1998)					

14. Taylor	Meta-	Not reported	Meta-analytical comparison of situational and	1. Validity coefficients:	Design of the	А
(2002)	analysis		past behaviour employment interview	Past behaviour: r = .63	included studies is	(diff)
			questions.	Situational: r = .47	unclear.	
	k = 30					
	(situational		1. Studies using past behaviour questions,	2. Past behaviour: r =		
	studies)		when used with descriptively anchored answer	.56		
			rating scales, yielded a substantially higher	Situational: r = .45		
	k = 19		mean validity estimate than studies using the			
	(past		situational question format with descriptively	3. Past behaviour, low:		
	behaviour		anchored answer rating scales.	r = .71		
	studies)			Past behaviour, high: r		
			2. Question type (situational versus past	= .56		
			behaviour) was found to moderate interview	Situational: ns		
			validity, after controlling for whether studies			
			used answer rating scales.			
			3. The validity estimate for past behaviour			
			questions used with			
			low-complexity jobs was higher than for more			
			complex jobs.			
			- Situational = ''Assume that you were faced			
			with the following situation what would you			
			do?''			
			- Past behaviour = ''Can you think of a time			
			when what did you do?''			

15. Whetzel	Meta-	Female and	1. On average, White test takers perform	1.	Design of the	А
(2008)	analysis	male, White,	better on Situational Judgment Tests than	Black– White: d = .38	included studies is	(diff)
		Black,	Black, Hispanic and Asian test takers.	Hispanic– White: d =	unclear.	
	k = 62	Hispanic and		.24		
		Asian	2. Female test takers perform slightly better	Asian– White: d = .29	Most effect sizes	
		situational	than male test takers on Situational		reported concern	
		judgment	Judgement Tests.	2.	vector	
		test takers.		Male– Female: d =11	correlations.	

Excluded studies

1st Author and year	Design	Reason for exclusion
1. Bobko (2013)	Literature review	The procedure/methodology of the review is unclear.
2. Hoffman (2015)	Meta-analysis	Too specific, not relevant to the review question.
3. Konradt (2020)	Meta-analysis	Too specific, does not address barriers or biases in the selection procedure.
4. Levashina (2014)	Meta-analysis	The procedure/methodology of the review is unclear. Design of the included studies is unclear.

5. Lievens (2017)	Non-systematic review	The procedure/methodology of the review is unclear.
6. Lievens (2021)	Non-systematic review	Focuses mainly on predictors for future performance, no effect sizes were pooled. In addition, the procedure/methodology of the review is unclear.
7. Rothstein (2006)	Non-systematic review	The procedure/methodology of the review is unclear.
8. Salgado (2019)	Meta-analysis	Only studies conducted by the U.S. Employment Service were included. GMA validity estimates were only slightly smaller than those obtained by Hunter and Hunter (1984).
9. Wiesner (1988)	Meta-analysis	Not up-to-date. Not relevant to the review question.

Appendix IV (papers)

Data extraction form - employee recruitment and selection - marginalised youth

Study level is rated in relation to its intended purpose (see p.5), i.e. to assess: effect or impact (eff); frequency (freq); or difference (diff).

1st Author and year	Design and sample size	Sector / Population	Main findings	Effect sizes	Limitations	Level
1. Buchanan (2016)	Qualitative study (observation and semi- structured interviews supported by a focus group) N = 36/15	young people (16-19) not in education, employment or training (NEET) in Scotland	 Disadvantaged youth face heightened barriers to information access and use beyond or not common to the generaladolescent population (e.g. low levels of literacy and self-efficacy). Disadvantaged youth face internalised behaviour barriers to information access and use beyond or not common to the general adolescent population. There is evidence suggestive of deception, risk-taking, secrecy and situational relevance in their (often self- protective) information behaviours. These findings are indicative of an impoverished information world. 	n.a.	Purposive sample, small sample size, serious risk of bias.	n.a

2. Butt (2010)	Randomised controlled study N = 1,600	adolescents from different states of Malaysia	The study explores the possibility of using ethnicity as a target variable for adolescents. 1. The target ethnic group of advertisements that use ethnic marketing had a more positive attitude towards the ad, the brand in the advertisement and a higher purchase intention compared with the non-target ethnic group.	not reported, (but the partial eta square indicates that ethnicity predicted 10 percent of the variation in dependent variables)	Unclear whether the findings are generalisable to other ethnic groups and other countries.	A (eff)
			2. The majority ethnic group had a more positive attitude towards the ad, the brand in the advertisement and a higher purchase intention for target advertisements compared with non-target advertisements.			
			3. Minority ethnic groups did not differ in their attitude towards the ad, the brand in the advertisement and purchase intention for target advertisements compared with non- target advertisements.			

			* Ethnic marketing = incorporating cultural and ethnic cues such as language, using an ad model of similar ethnic background, symbols or any combination of these in the communication strategy.			
3. Butt	Randomised	adolescents	1. Compared with non-targeted	not reported	Unclear whether the	А
(2012)	controlled study	from different states of Malaysia	respondents, targeted advertisement respondents had more positive attitudes towards the		findings are generalisable to other ethnic groups and other	(eff)
	N = 800		advertisement and the brand represented as well as greater subsequent purchasing intent.		countries.	
	(replication of Butt, 2010)					
			2. Interestingly, no significant differences were found between strong and weak ethnic identifiers in terms of attitudes towards targeted/non-targeted advertisements.			

4. Chapple (2002)	Qualitative study (interviews) N = 90	US (Bay area) adult mothers on welfare who were working part-time or not at all	1. For chronically unemployed women, informal job search methods in part explain poor labour market outcomes, whereas career- oriented women use network resources and education to connect to career paths.	n.a.	n.a.	n.a
			2. The networks of chronically unemployed women not only fail to lead to consistent or promising jobs, but the very existence of these networks as a resource enables them to avoid the labour market. In essence, informal job search keeps them in a self-perpetuating loop of connectedness to a secondary labour market – and unemployment.			
			3. A system is needed for connecting job seekers to employers through employment brokering programmes. These services, along with apprenticeship and mentoring programmes, are essentially strategies that help fill gaps in social networks, particularly for the chronically unemployed.			

5. Collins (2001)	Mixed methods (survey and focus groups) N = 764	US adult education students (South Carolina)	Study focuses on ways to communicate pro-social messages to overlooked and underserved societal subgroups.	n.a	n.a.
			1. Results suggest that traditional mass media may not be the most appropriate or efficient information channels for public relations and other communicators wishing to convey such pro-social messages to similar audiences (e.g. 68% indicated that either friends or a family member were the best options to reach others like themselves).		
			2. If mass media are employed, it may be necessary to rethink both the content and the intended receivers of such messages.		

6. Easley (2018)	Cross-sectional study	data drawn from several US Census data sets (2010)	1. Residential segregation from Whites is a central predictor of exposure to spatial mismatch across all minority groups, though findings suggest that this relationship is not driven by suburbanisation.	spatial mismatch Black β = .55, R ² = .64 Mexican β = .57, R ² = .51 Puerto Rican β = .45, R ² = .54 Cuban β = .45, R ² = .40 Other Hisp β = .46, R ² = .59 Chinese β = .44, R ² = .51 Vietnamese β = .55, R ² = .56 Korean β = .55, R ² = .57 Japanese β = .38, R ² = .44 Other Asian β =	Number of observations unclear.	A (diff)
				Other Asian β = .45, R² = .40		

7. Escott (2012)	Mixed methods (statistical analysis of labour market	young women living in disadvantaged communities	 Qualifications and ethnicity shape young women's relationships with the labour market. 	n.a.	No serious limitations.	A/ n.a. (diff)
	data, structured interviews, n = 80, and focus groups)	across England	2. In poorer communities, caring responsibilities and ill health are particularly important contributors to low levels of engagement.			(diff)
			3. Most young women have strong work aspirations and appropriate qualifications but considerable con- straints limit their horizons, in turn affecting their health and wellbeing.			
			4. Labour market vulnerability among young women is not simply the result of social characteristics > several other processes tend to reproduce young people's disengagement from the labour market. For example, the jobs available to the employable young women are often poor quality, low status, low pay and in a very narrow range of sectors and occupations.			
			5. Labour market activity was further damaged by age and race discrimination from local employers.			
			6. The failure of job search services was found to fuel the problem. Although there were exceptions, many job agencies did not appear to be grounded in an			

understanding of young women's specific circumstances. 7. Proximity to work, transport costs and social networks were repeatedly raised in the focus groups and young women often articulated the desire to work close to home, suggesting that for disadvantaged young people the nature of the <i>local</i> labour market is more important than for others, who are more likely to access the wider	
national labour market.	

8. Gordon (2006)	Systematic review, includes RCTs K = 35	mixed: young people, adults, minority ethnic and disadvantaged groups	1. Research illustrates that social marketing interventions of different types can be effective and that they can work with different target groups: young people, adults, minority ethnic and disadvantaged groups.	not reported	Included studies focused on smoking cessation, physical activity interventions, improving diet and substance misuse, so, hard to generalise to the REA target group.	AA (eff)
			2. Furthermore, the results demonstrated that social marketing interventions can be effective in a range of different settings:			
			schools, the workplace, church based, community and family- based settings, clinical practices, supermarkets and media-based settings amongst others.			
			3. The reviews also produced evidence that both narrow and broad focus social marketing interventions can be effective.			
			* Social marketing takes learning from the commercial sector and			
			applies it to the resolution of social and health problems. It focuses on voluntary behaviour change rather than coercion or enforcement.			

9. Klinthall (2016)	Cross-sectional study N = 605	Swedish-born young adults in Stockholm, whose parents were born either in Turkey or in Sweden	1. Results show that young persons whose parents are born in Turkey (Stratum T) use informal contacts in order to find employment more often than young persons whose parents are born in Sweden (Stratum S). Living in immigrant- dense areas increases the likelihood of finding employment through informal contacts.	OR's vary from 1.5 to 1.7	No serious limitations.	A (diff)
			2. For Stratum T, co-ethnic contacts are more important than other contacts, in particular if they are neighbours. Access to ethnic networks and the use of ethnic contacts in the labour market differ between persons of Turkish, Kurdish and Assyrian/Syriac backgrounds.			
			3. These findings suggest that young adults in immigrant-dense neighbourhoods have access to and benefit more from informal channels in order to find ways into the labour market.			

10. Perkic (2019)	Qualitative research (semi-structured interviews)	people responsible for the recruitment of disadvantaged youth in (Dutch) organisations	1. Results showed that almost three- quarters of the organisations recruit young people through their own social network.	n.a.	Master thesis, but methodological quality appears to be sufficient to include.	n.a.
	N = 16		2. Other methods, which cost more time and money, were mentioned to be used mainly when their own network was not (yet) built.			
			3. Organisations were able to reach disadvantaged youth because they actively targeted them in all their recruitment methods.			

11. Thornley (2010)	Systematic review of 45 papers covering 15 campaigns from the US, Australia, New Zealand and Tonga	Findings suggest that effective youth social marketing campaigns incorporated the following key features: 1. By youth for youth, where young people were included in all aspects of the campaign and their guidance was taken on board and used to shape and refine the campaign.	n.a.	Merely narrative and descriptive review. Methodological design and quality of the included papers unclear.	n.a
		 2. Use of methods and channels that were accessible and appealing to youth. 3. Ethnic-specific approaches. 			
		4. Targeting to subgroups such as adolescents e.g. 9–14 years and high-risk adolescents.			
		5. Aim for high exposure to a variety of different marketing activities.			
		6. Messages that empower youth and appeal to their need for independence and rebellion.			

			7. Strong and intense emotional messages (both positive and negative), use of social threat and personal testimony.			
12. Wozniak (2011)	Cross-sectional study N = 31	employers in the low-skilled labour market in South Bend, Indiana	1. Results provide a strong indication of racial bias during the recruitment and selection process: Employers believed young black male applicants were less likely to have the desired interpersonal skills and work ethic.	not reported	Limited information on research methodology, data analysis, etc.	A (freq)

Excluded studies

1st Author and year	Reason for exclusion
1. Aldridge (2008)	Outcome is depression or stress-related growth.
2. Keep (2012)	Not an empirical study.
3. Nwanko (1998)	Not an empirical study.
4. O'Donnel (2005)	Not an empirical study, narrative summary of findings/recommendations, focuses solely on recruiting disadvantaged youth for IT education and training.
5. Wentling (2001)	Qualitative study, focuses mainly on barriers that hinder the transition of minority youth in the workplace.