Factors influencing the completion of the Mountain Leader qualification

Will A. S. Hardy

2020-05-28

Contents

Thesis Abstract **9**

1 General Introduction **11**

2 Qualitative **13**

2.1 Introduction 13

2.2 Methods 13

2.2.1 Philosophical orientation 13

2.2.2 Participants 13

2.2.3 Interview guide 15

2.2.4 Credibility and trustworthiness 15

2.3 Results 16

2.3.1 Getting to Assessment 16

2.3.2 Motivation 19

2.3.3 Passing 30

2.3.4 Reassessment 30

2.4 Discussion 32

2.5 Conclusions 32

3 Key Discrimantory Factors **33**

3.1 Introduction 33

3.1.1 The Mountain Leader qualification 33

3.1.2 Pathways 35

3.1.3 This paper 36

3.2 Study 3 - Identifying the most important factors 36

3.2.1 Method 36

3.2.2 Results . . . 42
3.2.3 Discussion . . . 46
3.3 Discussion . . . 56
3.4 Future . . . 56
3.5 Conclusion . . . 57
3.6 Study 2 supplementary information . . . 57
3.6.1 Variables . . . 57
3.6.2 Study 2 classificaiton rates . . . 57
3.7 Study 3 supplementary information . . . 57
3.7.1 Example ominbus feature subsets . . . 57
3.7.2 Male getting to assessment training model performance . 62
3.7.3 Female getting to assessment training model performance 62
3.7.4 First time pass model performance . . . 62

4 Self-Efficacy and Quality Mountain Days 63

5 General Discussion 65

A Developing the survey tool 67
A.1 Study 1 - Item selection . . . 67
A.1.1 Method . . . 68
A.1.2 Results . . . 73
A.1.3 Discussion . . . 74
A.2 Study 2 - Item reduction . . . 74
A.2.1 Method . . . 74
A.2.2 Results . . . 80
A.2.3 Discussion . . . 81

List of Tables

3.1 Participant descriptive statistics. 37

3.2 Candidates pathway progress when completing the survey. 37

3.3 Group 5 male candidates getting to assessment within 18 months of training, unstandardised group descriptive statistics 42

3.4 Group 5 male candidates getting to assessment within 18 months of training, prediction model performance (n = 164). 44

3.5 Group 5 male candidates getting to assessment within 18 months of training, prediction model performance 44

3.6 Group 5 female candidates getting to assessment within 18 months of training, unstandardised group descriptive statistics . 46

3.7 Survey variables 58

3.8 Classification rates for the feature subset with the highest classification rates for each data set (percentage accuracy). 59

3.9 Group 5 male candidates getting to assessment within 18 months of training, training model performance 60

3.10 Group 5 female candidates getting to assessment within 18 months of training, training model performance 61

3.11 Group 5 passing first time, training model performance. Data standardised within sex 62

A.1 Latent variable correlations between full- and short-form measures. 73

A.2 Survey participants per group 76

List of Figures

3.1 Pass rates for female and male candidates assessed since 2000 (n = 13828). 34

3.2 Survival rates for female and male candidates post-training.Candidates trained 2009-2019. (n = 15635) Blue dashed line at mean time to assessment. 35

3.3 Study 3 participants. For simplicity, candidates who have not been assessed have not been added to this figure as a final group, therefore it can be assumed that candidates not progressing from one node to another have not been assessed. 39

3.4 Normalised training group means for male candidates getting to assessment within 18 months of their training course. 43

3.5 Normalised training group means for female candidates getting to assessment within 18 months of their training course. 45

3.6 Normalised training group means for candidates passing their first assessment - survey variables. 47

3.7 Normalised training group means for candidates passing their first assessment - DLOG variables. 48

A.1 Survey tool overlap 75

A.2 Study 2 participants. 77

Thesis Abstract

Chapter 1

General Introduction

Chapter 2

Qualitative

2.1 Introduction

2.2 Methods

2.2.1 Philosophical orientation

Who knows!

2.2.2 Participants

- A purposive, criterion sampling approach
  - What were my criteria?
- All white and >38 – is this important? Something about their experiences and viewpoints being similar/lacking the view of a younger/newer director
- Sampling criteria – purposive, gatekeepers. Became clear that I needed to interview some course staff as they have more interactions with candidates
- After gaining ethical approval from Bangor University's School of Sport, Health, and Exercise Sciences ethics committee, and individual informed consent, eight individuals participated in this study.
- Initial interviews carried out with five members of staff from Mountain Training, three of whom were the executive officers for: Mountain Training Cymru, England, and Scotland, the three main national training boards (four men and one woman, Mage = 47.19, SD = 6.60, MML courses = 41.6, SD = 29.86). Having completed these interviews it became clear that it would be important to interview course staff who had a greater knowledge of candidates and their experiences than the Mountain Training staff had. Therefore, we interviewed two high volume course providers and a course director who had worked for eleven different providers over 14 years (two men and one woman, Mage = 55.30, SD = 5.18, MML courses = 284.00, SD = 214.68).

- In line with University ethical guidelines, interviews commenced only when the participants had provided informed consent either through written or recorded word depending on level of visual impairment. This data collection process was chosen because when done well, interviewing can generate rich, retrospective, and thick descriptions of human life. Thus, interviews are a useful source of knowledge about personal and social aspects of lives. An interview guide was used to help facilitate discussion. Questions included in 'the guide were, 'Why did you decide to be physically active?,''Why have you not been physically active , 'What are some barriers to your physical activity? , 'How does being active' impact on your wellbeing? and 'How have you remained active? Clarification, elaboration, and detail orientated probes, that is, curiosity-driven follow-up questions were used throughout to elicit richer data.20 Participants were interviewed in a location of their choosing. This was primarily in their homes and occasionally in a local coffee shop or quiet meeting place. Interviews were digitally recorded, and all data were subsequently transcribed verbatim. Data were analysed using an inductive thematic analysis. This involved the systematic organisation, description, and interpretation of the key patterns (themes) within the data set. In terms of validity, the study was guided by a list of criteria' as assembled in Tracy's review of what constitutes 'excellent qualitative work (p. 837). This included: the worthiness of the topic; the significant contribution of the work; rich rigor, that is, developing a sample appropriate for the purpose of the study through the processes noted above and generating data that could provide for meaningful and significant claims; the coherence of the research, which refers to how well the study' conforms to the purpose, methods, and results as judged by an independent researcher and readers; 'member checks' by verbally sharing interpretations of data and inviting participants to reflect on these interpretations. Finally, an audit trail was used that involved a researcher independently scrutinised ethics, data collection and analysis. To that end, the researcher doing the majority of the data collection (M. Griffin) practiced conscious reflexivity, kept a reflexive journal and consulted with critical friends throughout the process. The project also included an independent advisory board that scrutinised ethics, the collection of data, analytic ideas, and practical recommendations that followed. The board comprised of five researchers who have published extensively on ageing, disability or physical activity, three representatives of relevant non-profit organisations, and two older adults with visual impairment.(Phoenix, Griffin, & Smith, 2015)

2.2.3 Interview guide

- How did I develop it?
  - Literature review, list of >50 factors, expectancy value exercise to rate each factor in terms of both ease of collection from interview and also usefulness, based on this 35(?) factors were included in the interview guide
  - Developed a list of prompts and probes to get the most out of people in an effort to reach data saturation
- What sort of questions?
- 5 Pilot interviews – 2 pre-pilots and 3 real ones. Data collection and analysis
- All interviews were carried out face-to-face in a location chosen by the participants (e.g., home or office space). Given the exploratory nature of the interviews and range of factors included in the interview schedule the interviews were carried out over multiple sessions to avoid fatigue of the participants and interviewer (Mduration = 316.25, SD = 54.85 mins, range: 2-5 sessions).
- Notes made during the interview
- Audio recorded and then transcribed clean verbatim by someone else due to volume
- Coded in NVivo 11 Pro using a concurrent inductive and deductive thematic analysis process. A priori deductive codes were created based on the literature and inductive codes were created as new themes emerged from the data. This combination of analytical procedures makes use of relevant literature but allows other factors to emerge from the data
  - The concurrent application of inductive and deductive methods contributes to the development of rich rigor by recognizing the scope and context of previous research literature, whilst also allowing for the identification of additional constructs of interest. This, in combination with the collection of rich and abundant data ensured that the complexity and nuances of the data were not missed. (Webster et al., 2017)
- Memoing during the analysis of each transcript – by participant and some codes
- Collection of data from different sources. MT executives, high volume providers, and course directors

2.2.4 Credibility and trustworthiness

- Member reflections. Due to the length of the transcripts (Mwords = 44907.88, SD = 7939.72) we returned an individual summary document to participants along with their transcripts that we asked them to check for authenticity. In addition to asking participants to check these documents for authenticity, we used two rounds of member reflections to better understand the interview data as well as generate new data (Smith

& McGannon, 2017; Tracy, 2010). In the first round we asked participants to reflect on their individual summaries and in the second round we asked them to comment on initial findings from all of the interviews. The data generated from these member reflections were included in the final analysis. Member reflections are not intended to find agreement but, as traditional member checking did, but instead they are intended to generate additional data and encourage reflexivity.

- Critical friend. WH presented themes and relevant excerpts from the data to RR, who offered critical feedback. The aim of this was not to reach agreement or consensus but to encourage reflexivity (Smith and McGannon, 2018).
- The first author, who has 10 years of outdoor experience across the world, conducted all of the interviews. Most of this experience has been gained in a recreational context rather than a professional one, this is seen as a strength because the analysis was less likely to be influenced by personal experiences of Mountain Training qualifications, ensuring that it is the participants experiences that are presented. In addition to this, RR is an SL in Sports Psychology with XX years of outdoor experience and LH is a Professor in Sports Psychology and has XX years of outdoor experience and is an IFMGA Mountain Guide. The experiences of the research team meant that good rapports could be established with participants and that the subtleties of the phenomena of interest could be fully understood.
- The first author was instructed in qualitative research methods by the second and third authors and additional knowledge was gleaned from recent literature on qualitative research methods
- Research team experience of qualitative research

2.3 Results

2.3.1 Getting to Assessment

Three main themes emerged from the interviews as factors that influenced candidates' likelihood of being assessed: confidence, motivation, and gaining experience. There were three additional themes that participants felt influenced whether candidates reached assessment (albeit to a lesser degree): candidate location, reengaging later in life, and redirection to lower qualifications.

2.3.1.1 Confidence

The results in this section shows that candidates must be confident in their ability to pass a five-day Mountain Leader assessment before they will attend one and the threshold of confidence required is individual to them. There are several factors that influence both the level of confidence and the individualised thresholds that candidates must surpass. Level of confidence.

2.3.1.1.1 Level of confidence.

All seven participants said that candidates needed to feel confident before they would attend an assessment. O1 said:

> [Candidates] have to put themselves on a little pedestal and go, "This is me, and this is how I'm trying to go through the scheme." … That takes someone who's got a reasonable amount of confidence in themselves to do that. I can imagine some candidates not feeling comfortable in putting themselves in that position … and I think that they will be the ones less likely to complete.

O2 supported this when describing candidates on assessment courses by saying, "In their heads, they're prepared for it." suggesting that only well-prepared, and therefore confident, candidates attended assessment courses. Talking about candidates who did not attend assessments P1 said, "They convince themselves they're not ready, and then they won't book on." However, O1 suggested that more than just experience was needed for candidates to feel confident when he talked about a candidate who, "Doesn't have the confidence to do the assessment" despite them being a "Super keen hillwalker … who has done the training."

This evidence shows that the candidates who have reached assessment were confident in their abilities and that some of those who have not reached assessment did not feel confident, it also suggests that their confidence was not always dependant on their abilities.

2.3.1.1.1.1 Individual differences in thresholds of confidence.

This section presents evidence that candidates have their own thresholds for confidence that they must surpass before they will attend an assessment together with factors that influence that threshold, thus moderating the relationship between level of confidence and the likelihood of booking an assessment.

Age.

Five participants suggested that younger candidates have lower thresholds for confidence and that older candidates were less likely to feel confident enough to attend an assessment. O4 said, "younger folk can be less constrained by lack of confidence." P2 supported this, suggesting that if older candidates did not feel confident they are more likely to refrain from booking an assessment, "Some of the older guys and girls have come in already with 40 days but they still might not come back for a year or two because they're sometimes not as confident." O1 and O2 did not comment on how age may or may not relate to confidence and getting to assessment.

Gender.

Five participants discussed the effect of gender on confidence and all said that females needed to be more confident than males to attend an assessment. When asked if there were many candidates who were ready for their assessment but did not feel ready, and so did not attend an assessment, P1 said: I think a lot of girls fall into that category. That they actually could do it, but it feels like such a big thing. They want every "i" dotted and every "t" crossed, and they want to be absolutely doubly sure that they can do it, and really, they could have done it earlier.

Perfectionism.

Five participants suggested that some candidates may not have attended an assessment because their perfectionistic traits led to them having a higher threshold for confidence, thus not feeling confident enough to attend and assessment. P2 gave an example where high-levels perfectionistic strivings may have led to candidates not feeling ready for an assessment despite being ready, "For some reason or another, they've really held back … it could be that they're an absolute perfectionist and they just didn't want to turn up until they were totally happy." O3 suggested that female candidates had higher levels of perfectionistic concerns, thus were more likely to incorrectly feel that they were "below the standard", "A female might actually be overly cautious about exposing themselves, and potentially failing … through believing they're actually below the standard. Whereas they're probably higher than that." O1 and P3 did not discuss how perfectionism may or may not influence candidates' confidence threshold.

2.3.1.1.1.2 Understanding the standard.

Throughout the interviews all participants referenced "the standard" (i.e., the standard required to pass) and commented that it is often hard for candidates, and sometimes staff, to understand what it is. The five participants that discussed "the standard" and how it related to getting to assessment all suggested that a candidate's confidence level may not surpass the threshold needed to attend an assessment because they did not understand "the standard", thus making it hard to be confident. P3 explained that this holds some candidates back from being assessed: They need that reassurance that … they're consolidating correctly, and actually they're performing at the standard … because they're not going to come forward unless they feel like that …. I think that's really hard [for candidates] to know where they're at in relation to the bar. We think it's clear … but candidates always ask, "How close do I have to be? You know, like, ten metres, a hundred metres. One contour line, two contour lines." Three other participants made similar comments, and O2 did not refer to understanding the standard.

2.3.1.1.1.3 Raising candidates' confidence levels.

Six participants discussed how support helped close the gap between candidates' confidence levels and their confidence thresholds by raising confidence levels rather than lowering confidence thresholds. When talking about candidates who lacked confidence P3 said: They often need a lot more support, and with a bit of support they often shine as well: as soon as they realise that, "Actually, I am good enough and I can do it", then they're up and running, although it can be fragile, and it doesn't take much to knock it. Talking about isolated candidates O4 said, "I suppose the ones without a network … and those at the lower end of the confidence spectrum … are going to need help with upskilling or believing that they've got the skills in the first place."

2.3.1.1.2 Gender differences in robustness of confidence.

In addition to the gender differences in confidence thresholds discussed above, three participants spoke about gender differences in the robustness of candidates' confidence. When talking about the different influences of negative events on candidates, O3 said, "Who can take it in their stride? Blokes, I suppose. Not because they perform well afterwards, they will probably be weaker. They are more Bolshie, I suppose." O2 supported this:

> For some candidates, particularly men, those effects of that bit of negative feedback or that bad day they had on the hill, they try and brush off and just carry on … and get it right next time …. What you might find with many females is that's thrown a spanner in the works. It's made them doubt what they need to do, and now they need to readjust their consolidation plan.

2.3.2 Motivation

Many motivation researchers have proposed that motivation operates at different levels (e.g., Vallerand, 1997; Vallerand & Blssonnette, 1992). In particular, self-determination theorists have proposed three levels of motivation: dispositional motives (i.e., goals for life in general), participatory motives (i.e., what someone hopes to achieve or avoid by participating in a behaviour), and regulatory motives (i.e., the perceived loci of causality of behavioural goals) (Deci & Ryan, 2000; Ingledew, Markland, & Ferguson, 2009). All seven participants gave examples of candidates with different participatory motives who also had different regulatory motives within those participatory motives. They suggested that both levels of motive influence candidates' likelihood of attending assessment.

2.3.2.0.1 Participatory motives.

All participants said that candidates with extrinsic participatory motives (i.e., to achieve an external goal), particularly allowing them to work, are more likely to complete than those with participatory intrinsic motives (i.e., doing something for its own sake). P1 said "The ones where there's a driver, are more likely to [complete] …. If they're not doing it for work and they're using it in an informal thing, they are probably less likely to [complete]." P1 went on to say, "People who want to use it for their work: formally or informally, directed or volunteer … they're pretty motivated to do it, and so I would say I think that the success rates are pretty good." Similarly, O3 said:

> If there's an expectation that somebody's going to have their ML to be able to do their job … I should imagine they get on with it. But if there's no real drive … [they] kick it down the road and, "I'll get around to it, maybe, or maybe not. It's not a big deal." sort of thing.

Five participants said that some candidates had registered for the ML to develop their personal skills and that for some of these candidates passing an assessment was not important, O4 said, “The ones doing it for their own skill improvement, it’s not part of a definite plan … they’re not so concerned if they complete or when they complete the award.” However, O3 did not believe that candidates attend a training course without any intention of going onto assessment but did think that some will decide not to continue:

> I don’t transpire [sic] to this “doing the mountain leader training course for a skills course”, to up-skill for an individual …. I can see how people would do it to start with, thinking they were going to progress to assessment, work out what were the demands upon them of attending an assessment, decide that we’re going to call it a day there.

P2 did not talk about candidates who only registered for the ML to develop their personal skills.

2.3.2.0.2 Regulatory motives.

Regulatory motives can be placed on a continuum from autonomous to controlled. Intrinsic-, integrated-, and identified regulatory motives are examples of autonomous regulatory motives, where behaviour is self-determined behaviour as the value of it is internalised. Whereas controlled regulation includes introjected- and external regulation, where behaviour is nonself-determined and the value of it may only be slightly internalised or not at all, thus controlled by external factors (Deci & Ryan, 2000). In these data, participants gave examples of candidates who had different regulatory motives and the influence that these had on candidates’ likelihood of attending an assessment.

2.3.2.0.2.1 Autonomous regulatory motives.

All seven participants said that those candidates who wanted to be outdoor instructors got to assessment. O1 said, “[If] they’re wanting to work in the outdoor sector they will naturally get [to assessment].” P3 suggested that those with autonomous regulatory motives were more likely to get to assessment, “If you’ve got people that are thinking about a full-time career in the outdoors … they are going to be more inclined to follow the process through.”

Another example of candidates having different types of participatory- and regulatory motives was seen in candidates who aspired to hold higher Mountain Training qualifications, of which the ML is a prerequisite for. O2 said, “they’ll tell you, ‘I am doing this because I want to do my MIA.’ ” Participants suggested that these candidates were extrinsically motivated but had autonomous behavioural motives. O3 gave supported this when describing his own experience of becoming a Mountain Leader, “I didn’t even want to do my ML, I just wanted to go and be an MIA. I was only interested in that …. I was pretty flipping motivated to get through this thing as fast as I could.”

O4 suggested that candidates who aspired to hold higher Mountain Training qualifications wanted to complete the ML quickly to progress, “Folk that have got a definite plan for using their ML, like they want to become an IML or whatever either will pursue it in a shorter time frame."

2.3.2.0.2.2 Controlled regulatory motives.

All seven participants talked about candidates who had controlled regulatory motives and suggested that these candidates were less likely to be assessed than those with more autonomous regulatory motives. When talking about which candidates attend assessments O1 said, "If the school has sent them there because they're going to run a Duke of Edinburgh, then no. They won't do it." P3 supported this and said that is because these candidates had not gained the necessary experience:

> We see a lot of people coming through with Duke of Edinburgh and Scouts who I'd say are pushed into it … they don't have the experience – the mountain experience as opposed to, sort of, hill and moorland experience – and it can be a shock. And then actually progressing through to assessment: they sort of realise, "Hang on." Yes, "I can't do this," or, "This isn't for me."

2.3.2.0.2.3 Intrinsic regulatory motives.

Candidates with intrinsic regulatory motives also had intrinsic participatory motives, at least to attend training. Those who did not feel that they wanted to be assessed were intrinsically motivated to attend a training course but amotivated to complete the qualification. P1 suggested that if candidates registered for the ML to develop their personal skills and found their training course inspiring then they were more likely to want to be assessed:

> If you run a good course, you enthuse them so much that there's no requirement on them to come back and do the assessment, but they actually want to do the assessment because they feel that it's a good challenge for their hobby.

This was supported by P3:

> Quite a few who come on training courses and say, "Oh, I'm just doing this for a personal thing," actually really enjoy it, and then they go, "Oh, I'm going to carry on now and do the assessment, and actually this seems like a really cool thing."

O4 suggested that candidates' motivation and self-efficacy can be influence by course staff:

> It's a combination, isn't it? Of helping them believe they can do it and helping them want to do it, to see value in completing, because a lot of folk come on training courses not being sure they need to do the assessment.

2.3.2.0.3 Negative disconfirmatory experience.

All participants talked about disconfirmatory experiences that reduced candidates' motivation to attend assessments. However, three of the participants also provided evidence that not all candidates who have these experiences will drop out. O2 proposed that all candidates will have at least one such experience, "I would be really surprised if they have never had a disconfirmatory experience." O1 and O3 go further and suggest that some candidates may become more motivated following a negative disconfirmatory experience. O1 summarised the possible effects of negative events on getting to assessment by saying that, "[candidates] either do a U-turn and don't bother or they up their game."

Five participants gave examples where candidates were part way through the ML process and realised that it was not something that they either needed to or could do. O1 gave the following example:

> Someone who … saw a Mountain Leader working, thought, "That's the thing for me," … and then once they started the process realised [that] actually there's a lot more to it than they were hoping and then become disinterested with how much experience they needed to gain from then on it, and then dropped off.

2.3.2.0.3.1 Negative experiences at training.

Six participants suggested that in some instances an ML Training Course itself could be a negative experience. When asked for an example of a disconfirmatory experience, O3 said, "[a disconfirmatory experience] might be just feeling they are well off the mark during a training course … that can be quite depressing … just not really nailing it on the training and then getting disillusioned."

When talking about candidates who felt less willing to attend an assessment P2 said, "People say, 'It really put me off. The training course really put me off,' and that's a shame when you hear that because they say, 'It was just awful.' " P2 repeated examples that candidates had previously given to them of reasons they had become less willing to attend an assessment:

> A lot of comments come, "Our training was worse than the assessment." … "We never had any feedback. We were assessed basically." … These people went on their training course and felt like they were beasted and battered and scrutinised like as if they were being assessed.

Six participants spoke about candidates who had not understood the purpose of the qualification when they registered for the ML and once they better understood the purpose of the qualification they realised/decided that they could/would not complete it. P3 explained that the training course had sometimes ben the stimulus for candidates making that choice, "We definitely get [candidates] that are coming forward and then they do the training course and they realise it is just not for them, they are not going to be able to put the time and effort in."

Some of the candidates who decided that they could/would not complete the qualification following their training course may have done so based on incorrect information. O1 said:

> We have had cases where someone has asked about experience [needed prior to assessment] and a provider has gone, "Well, in my view everyone needs to go to Scotland and go to the Highlands to gain experience." … Suddenly people are going, "Oh, my God. I live in the South East …. If I have to go to Scotland that's a whole different ballgame."

O1 went on to explain that the quality of information provided by training staff determined if it had a positive or negative influence on candidates, "The wrong kind of responses [from training staff] can have an impact. Whereas the right answers might mean that people get the correct information and can then plan accordingly."

O3 provided an example where candidates' perception of the course staff as role models might discourage them from completing, "I am sure there is nothing more disengaging than seeing somebody out of shape, out of currency doing a crap job on the hill. It is hard to engage with that."

2.3.2.0.4 Competing influences.

Five participants spoke about candidates who wanted to complete the ML but were not motivated enough to find the time to prepare for and then attend an assessment. There was some evidence that those who take longer to complete the ML will need more enduring motivation. O3 said, "Sometimes I think momentum is everything." O1 supported this:

> I think those who see it as the end goal take longer, and the more time that you put in between that training and assessment there are more variables of life that can get in the way that would then push that to the back burner.

When asked about candidates who were ready for assessment yet did not attend one, O3 said that the ML is, "an easy can to kick on down the street if you're busy with other parts of your life." O2 supported this idea of candidates having put their assessment off because they were busy with other things:

> Maybe they haven't turned up to assessment at that point because they haven't got the days, and said, "You know what, I haven't managed to get the days in, I'll leave it this year, I'll do it next year." That's fairly common …. There are just other things, life's got busy in other ways.

O1 explained that following a training course some candidates realised that they would need longer than previously expected to complete the ML. For some of those candidates their motivation to complete the ML did not last:

> Where candidates lose focus is if they've found that the training course has brought lots of new skills to them that they haven't seen before, they start pushing back when their assessment time's going to be. I think once that goes beyond 12 months, they kind of come off the boil with their consolidation time because it feels like there's no urgency …. I think once they do that they're less committed, so making good use of their free time to consolidate and gain further experience becomes less of a priority, so the further that goal is the less a priority it becomes in their everyday life. Then that opens up lots of opportunity for life events to get in the way.

2.3.2.1 Gaining experience

One prerequisite for a candidate to attend an assessment is having a minimum experience of 40 Quality Mountain Days (QMDs). Accruing 40 QMDs requires the investment of both time and money. All seven participants discussed reasons that candidates had not meet this prerequisite and thus did not attend assessments.

2.3.2.1.1 Barriers.

All seven participants spoke about aspects of candidates' lives that prevented them from gaining sufficient experience to get to assessment. O1 said, "If people can't get the experience they can't proceed." P3 supported this by saying, "Location and time, I would say are the biggest two handicaps for people. So, if you don't live in the mountains and you've got a fulltime job and a family, really hard." When asked how different motives for doing the ML influenced a candidates' chances of completion P1 said:

Well, really, it boils back to, "Are they in a position to gain that experience to go forward to assessment?" That's the actual crucial thing, I think, more so than any one group where you go, "Yes, they're much more likely to do it." O3 supported this saying, "I think timing is critical, you have got to have the time to gain experience. You have got to have enough money in the bank to get through the process."

2.3.2.1.1.1 Lack of time.

Participants gave three main reasons that candidates felt they lacked time to prepare for their ML assessment: profession, family, and doing other multiple qualifications at the same time. These other domains of candidates' lives became barriers to completion for them as they were more important to those candidates than becoming Mountain Leaders.

Profession.

All seven participants suggested that candidates whose profession allowed them time to prepare were more likely to be assessed than those whose profession did not. How a candidate's job is set up appears to be more important than what that job is.

An example of candidates in the same profession having different amounts of time to prepare is clearly illustrated amongst trainee instructors; five participants spoke about how different trainee instructor schemes influenced how much time candidates felt they had to prepare. When asked how being a trainee might affect a candidate's chances of completion P3 said, "[Outdoor Centre A] and people like that with, some of their staff are very good at giving them time off, or sometimes even paid time to go and do a bit of personal development." And when talking about candidates from outdoor-activity centres P2 said:

> If you're just given week after week of programmes that demand your time, working at low level, and the organisation is not giving time to develop their own skills … It's down to the company you're working for and it's down to the organisation. They're the ones who will decide what they need and how much time they've got available to release.

This was also evident in the five interviews where participants spoke about how being a teacher influenced a candidate's likelihood of attending an assessment. P1 explained that teachers who felt that they only had their holidays to prepare for the ML might have felt that they could not "fit it in" and that teachers available time is dependent on their school's view of the ML:

> I mean, schools can be helpful or not so helpful …. If the head teacher gets outdoor ed. and all the good things that spin out of it, then they can be very supportive. If the head teacher doesn't, then the teacher's kind of fighting them as well with all the other pressures: family, money and whatever.

Family.

All participants said that candidates having family commitments would make them feel that they had less time to prepare, so were less likely to get to assessment. For some candidates, this was moderated by support from their family, allowing candidates to prepare for the ML instead of fulfilling their family commitments. When asked for examples of reasons people have given for not completing the ML O2 said: Family. Family and work. Kids, or family circumstances, maybe elderly parents. That seems to be the main thing, or work commitments …. Sometimes they come back … they have resurfaced on the other side to say, "I am picking this back up again."

O1 gave an example of a candidate whose family situation, and thus priorities, changed between training and assessment, which meant that they had not and were unlikely to complete the ML:

> Three years ago, I talked to her about doing the ML. She cracks on with doing that. She's done the training. She hasn't done the assessment. She's now had a kid, and it's almost totally irrelevant to talk to her about ML these days.

O2 explained that candidates from different backgrounds will have different levels of family responsibility when talking about candidates from minority groups, “Sometimes when folk in other communities get involved in the outdoors there are religious, cultural and social pressures …. Family commitments come first, and it has a big impact on free time … suddenly your free time isn’t free.”

Multiple qualifications.

Some candidates also work towards other qualifications at the same time as the ML. Five participants suggested that working towards multiple qualifications at the same time has a negative impact on the time available to candidates and thus their likelihood of attending an assessment. O2 explained that working towards multiple qualifications at the same time made it harder to do one well:

> [Candidates] who tried to then spread with paddle sports and that really suffered …. You have to have a bit of a focus …. You have to decide which one it is you are going to do. Unless you are one of these really rare people who’s brilliant at everything.

P3 suggested that working towards multiple qualifications at the same time may be detrimental to a candidate’s chances of attending an assessment because of changes in their regulatory motives:

> Sometimes they’re trying to do quite a lot of tickets all at the same time and it can become a chore for them, and it’s almost like a hoop that they feel they need to jump through as opposed to actually enjoying the process … I think a lot of them find it really hard to put the time in.

O4 also recognised that working towards multiple qualifications at the same time may limit the amount of time that candidates can gain experience in but suggested that there might be some advantages to this as well:

> [Trainees] might be preparing for other things at the time. But equally, they’re in a particular phase of their life and mind-set, which is award focused. So, therefore, they will be quite good at preparing for assessments and more likely to have access to other people that have got MLs that can help them.

2.3.2.1.1.2 Location

Six participants discussed how the place where a candidate lives influences how easily they can accrue QMDs. It is harder for candidates who live further from the mountains to accrue QMDs as they must both travel for longer and often feel that they need to take block of time off to get to the mountains. O2 explained that candidates living in Scotland could gain QMDs “in a day rather than two days” because they did not “have a day’s travelling to get there and back.” This was supported by P1 who said:

> "People for whom the mountains are a long way away: by definition, it's going to be harder because they've got to have the time and the money to get themselves there." P1 went on to say, "They're going to do it more as bunches of days, so they're quite likely to do multi-day expeditions .... Whereas, the people who live closer can do it weekend and weekend, once a month on a Sunday."

In addition, candidates living further from the mountains will face a higher financial cost. For some candidates, this can seem beyond their means, O1 said:

> The financial cost of gaining the experience is a massive challenge. When you're talking to someone from the South East, telling them they need to get up into Snowdonia and The Lakes, or The Highlands, on 40 occasions, they start going, "Bloody hell. I can't afford that."

2.3.2.1.2 Social support.

2.3.2.1.2.1 Development plans.

All seven participants said that it was important for candidates to leave their assessment with an understanding of what they needed to do to prepare for an assessment (i.e., have a development plan). When asked what the most important part of support was for candidates, P2 said:

> Once they've got onto the training a really good training course, which makes it clear to the candidates what it's all about, and then directs them the right way. You need to individually debrief people and get to know what their personal needs are ... A generic debrief really sometimes doesn't cover it thoroughly enough for individuals.

When asked what influence they thought the post-training debrief has on candidates P3 suggested that it could have a profound impact on candidates' expectations:

> It's a really important chat, ... it's really common on a debrief when you sit down with somebody and say, "That was an awesome performance. All you need to do is pad this logbook a bit, and you could come forward for assessment really quickly." They sit there and go, "But I was thinking about doing it in four years' time." and you're like, "What? You could do it next spring, no problems at all." ... You can have a big impact.

However, O3 explained that providing individualised feedback can be at odds with preventing training courses feeling like an assessment, an issue highlighted above (see [Negative experiences at training]), "I don't believe that candidates should feel they're under any sort of assessment process while on the training course. Once you have a formalised one-to-one debrief it can feel like an assessment." P1 suggests that it is possible to provide individualised feedback without making candidates feel that they have been assessed:

> My debrief is actually getting them to tell me what they think they need to do rather than me telling them what they need to do, because I would've had to assess them somehow to do that …. I'm asking them to self-assess and tell me what they think they need to do to get to the assessment.

2.3.2.1.2.2 Time support.

As shown above (see [Profession] and [Family]) some candidates felt that they did not have enough time to prepare for an ML assessment. However, different candidates with the same demands on their time can feel differently about the amount of available time they have. One reason for this is that some candidates are supported by their employers and families. When asked what sort of support candidates might look for O2 said,

> Having the support of their family is going to be absolutely paramount …. Having support from family to free up time and then actually having the time both from family and work that coincides with the others …. It is an acknowledgement within the family that [the ML] is important to the person. The ones who have succeeded against the odds have had that support. That's been really obvious.

Employers are another source of time support for some candidates. When talking about support candidates received with practical matters, P1 said, "Some of them are in organisations and centres where the management are on the ball enough to allow them development time."

2.3.2.1.2.3 Financial support.

Six participants spoke about candidates who had received financial support. In some instances, this was essential to candidates' progression to assessment. P2 said that, "A lot of people wouldn't be able to do ML if they didn't get financial assistance" and went on to say, "However, participants also suggested that financial support will only benefit candidates if they are also sufficiently motivated to complete the ML." O3 said:

> In my experience, those [whose] pathway has been paid for or financially supported, they don't really seem to engage with the actual role of taking responsibility for a group in the mountains …. Heavily subsided or full payment I tend to find they don't get a good solid engagement and on occasions people just don't turn up because there's no engagement at all.

When talking about candidates who want to use the ML for work P1 supported this paradox associated with financial support, saying, “they’re pretty motivated to do it, and so I would say I think that the success rates are pretty good for that, particularly if they’ve paid for it.”

2.3.2.2 Reengaging later in life

Five participants discussed candidates who had disengaged with the ML and then but reengaged with it later in life. P2 gave an example where candidates had an enduring motivation to become Mountain Leaders but had not completed the qualification because they were busy with other aspects of their lives:

> They start the process when they were young, free and single. They meet somebody, get married, have kids, they don’t do it for years and years and years. Then they come back to it. It’s something they’ve always wanted to do.

P1 also suggested that changes in family circumstances can be the reason that candidates reengaged with the ML:

> The Scouts, the Guides and the D of E are often the kick-start to get people back into it again because they’ve suddenly found that their kids are actually at that stage …. Then, they want some formal training on top of that.

P3 suggested that retirement might also provide candidates with an opportunity to reengage, “[Candidates] who did their training a long, long time ago and then their career is coming to an end …. They’ll reengage as well.”

2.3.2.3 Redirection to lower qualifications

Five participants suggested that after ML training some candidates decided to pursue a lower qualification instead, O1 said, “They can’t put [the ML] as the priority in their life, so they may drop back to the Hill and Moorland Leader or the Lowland Leader course as a more achievable objective.” This was supported by O2 above (see [Family]) and when talking about candidates who have struggled with the Mountain Leader training course, “We get a reasonable number that then convert to Hill and Moorland Leader …. They decide that they are going to do that, because that is a shorter assessment and less intensive.”

It is unclear how this will ultimately influence getting to an ML assessment. For some candidates this lower qualification will suit their needs and they will not continue with the ML but for others, completing the lower qualification becomes another step in the process of becoming a Mountain Leader. When talking about training debriefs, P2 said, “Sometimes, we would advise somebody to go and do the Hill and Moorland assessment …. They worked really hard to get the Hill and Moorland … then eventually, after a couple of years, they’ve done the ML assessment.”

2.3.3 Passing

2.3.4 Reassessment

Candidates who do not pass their initial assessment may or may not return to be reassessed. Whilst the interview was not designed to answer questions about what factors influence if candidates return to be reassessed, the semi-structured nature of them meant that some data emerged that provides some insight into this. However, these results are not as clear as those in previous sections.

2.3.4.1 Understanding the Original Result

All seven participants spoke about candidates either understanding and accepting their original assessment result or not. Participants suggested that candidates who understand and accept their result are in a better position to decide if they want to continue with the ML and if they do, understand what they need to do to pass a reassessment.

2.3.4.1.1 Preparing for reassessment.

Four participants spoke about candidates who realised that they were below the standard and then went away to prepare for reassessment, P3 said:

> You'll get lots of candidates who get deferred on their navigation, and are like, "Urgh," then they go away, sort themselves out, come back for reassessment, and at the reassessment process they go, "I definitely wasn't good enough, and I've gone away and done all this stuff, I know realise I'm a much better navigator than I was before."

2.3.4.1.2 Disagree/don't understand result

Three participants spoke about candidates who did not either agree with or understand their original assessment result. O4 gives an example of why candidates might not agree with their result: I guess there's a danger that [candidates] don't fully understand, the reasons for having been deferred. If they're pinning it on isolated, you know, isolated mistakes that they've made or errors. Maybe they haven't grasped that it's a pattern that's emerged.

2.3.4.1.3 Consequences of not understanding/accepting the result.

O1 explained that candidates who felt that their result was unfair would do one of two things: They would literally finish that assessment. Get the result they didn't want to hear. Then they will either do one of two things, complain, or just get annoyed, and try and book onto the next earliest assessment they can.

They don't believe they need to retrain. They believe they need to just be assessed again. Then they go to that next assessment and, hey presto, the same result .... Because nothing has changed. Unless it is about the assessor/candidate relationship .... If it's about the system, rather than about the assessor, then if they go on to the next assessment they will just get the same result again.

2.3.4.1.4 Reasons for not understanding/agreeing.

P2 suggested that clashes between candidates and staff are not uncommon:

> You always have people complain about something or somebody, sometimes, about situations they're in. You get a lot of info when people are being reassessed because they'd been deferred so you run a re-assessment and because they're being deferred, they start telling you why they think they shouldn't have been deferred and then they start slagging off providers and organisation.

P2 goes on to explain that in some instances these clashes can be highly charged, "I have heard stories of people saying, 'I nearly punched him. I nearly hit him. In fact, we all did. We all felt like turning round and hitting him.' That's not good, is it?"

2.3.4.2 Nerves

Three participants said that candidates who present for reassessment are nervous, P2 said, "Everybody who turns up for a re-assessment is full of nerves. They're very nervous when they start". Two of these three participants suggest that for some this nervousness can be so extreme that it manifests itself with physical symptoms. P1 gave an example:

> Some people are literally sick with worry on ML assessments. I mean, I remember doing a reassessment for this one guy and he confessed afterwards that, just before we'd met up, he was throwing up because he was that nervous about doing this.

P1 supports this, suggesting that it is not a one-off occurrence:

> They are really nervous and when you meet them you've got to really make sure you calm them down and you've got to try to create a really relaxed atmosphere before you set off because they're shaking some of them. They're nearly sick.

This nervousness may be in part due to their experience on their original assessment. P2 provided the following insight in candidates' original experience of assessment, "[Candidates] felt like [assessors] were quite harsh and quite lacking in any form of feedback or lacking in any form of empathy, which made them feel very uncomfortable, which made their performance even worse"

2.3.4.3 Redirected towards a lower qualification

Some candidate will not return for a reassessment because following their initial assessment because the assessment staff have redirected them towards a lower level qualification, as the assessment staff feel that would be more appropriate for them. Whilst only P2 spoke about candidates being directed towards a lower level qualification rather than reassessment this redirection is also spoken about in Getting to Assessment. P2 said, "When you get people like that, we advise them to do Hill and Moorland…. You will get that candidate who will be better off, definitely, doing Hill and Moorland."

+++from what is here it seems like only understanding and accepting results (or not) is the thing that influences returning to reassessment, with this understanding and acceptance likely influenced by the relationship between candidates and assessors. There are other things in your diagram (reprepare for the standard) that are not really mentioned here. A lot of the diagram makes sense to me, but it doesn't feel like all the things from the diagram necessarily come through in the story. This might just be me though, we can discuss+++

Should this be returning and passing reassessment?

Make about reassessment in general rather than returning for reassessment.

2.4 Discussion

2.5 Conclusions

Chapter 3

Key Discrimantory Factors

3.1 Introduction

The findings reported in this paper are part of a larger, Mountain Training United Kingdom and Ireland (MTUKI) sponsored, project. This paper builds on a large qualitative study reported in Chapter 2, the aim of which was to identify factors that may influence the completion of the Mountain Leader qualification (ML). From the results of that study, it was clear that there was no single factor that influenced completion, but instead the results suggested that a myriad of factors are important and there are multiple interactions between them.

3.1.1 The Mountain Leader qualification

The ML is a qualification for "people who want to lead groups in the mountains, hills and moorlands of the UK and Ireland" (Mountain Training UK, 2015a). To become an ML, candidates must complete a training pathway, involving four distinct steps, which have remained broadly similar since 1954 when the qualification was created.

- Something about projects in other countries being based on the ML - UIAA w/Steve Long

To qualify, candidates must first *register for the qualification* and gain a minimum of 20 Quality Mountain Days (QMDs)[1], then they complete a six day *training course*, following that they are required to gain a minimum of 20 additional QMDS as *further experience to consolidate skills*, and finally they need to successfully complete a five day *assessment course*. Therefore, to become an ML, one must spend a minimum of 51 days in the mountains. Most successful candidates will have more experience than this, whether that is additional

[1]There is not a simple definition for a QMD, however it is expected that a QMD will "make a positive contribution towards a person's development and maturity as an all round mountaineer" (Mountain Training, 2019)

QMDs, experience of mountain walking that does not meet the QMD criteria, or other mountaineering experience. Becoming an ML requires candidates to commit a significant amount of time and money in order for them to qualify as MLs.

Between 2009 and 2018 an average of 2278 candidates registered for the ML qualification each year, but only 559 qualified. When looking more closely at the numbers of candidates who did qualify it became clear that there are two main components to qualifying: 1) getting to an assessment and 2) passing an assessment. Interestingly, most candidates who got to an assessment passed their first assessment (Figure 3.1), but most candidates did not get to an assessment (Figure 3.2). It is also noteworthy that, as shown in Figure 3.2, becoming an ML is not a quick process, the mean time from training to assessment was 1.57 years.

Mountain Training UKI would like to understand why people do not complete the ML qualification and identify if there are any changes to the pathway that they can make in order to better support their candidates. It is unlikely that there is a single factor that would be a "silver bullet" to improve completion rates. Instead there are likely a myriad of factors which influence completion at various stages of the pathway. Some of these factors will be generic to all candidates, whilst some may be specific to individual (groups of) candidates.

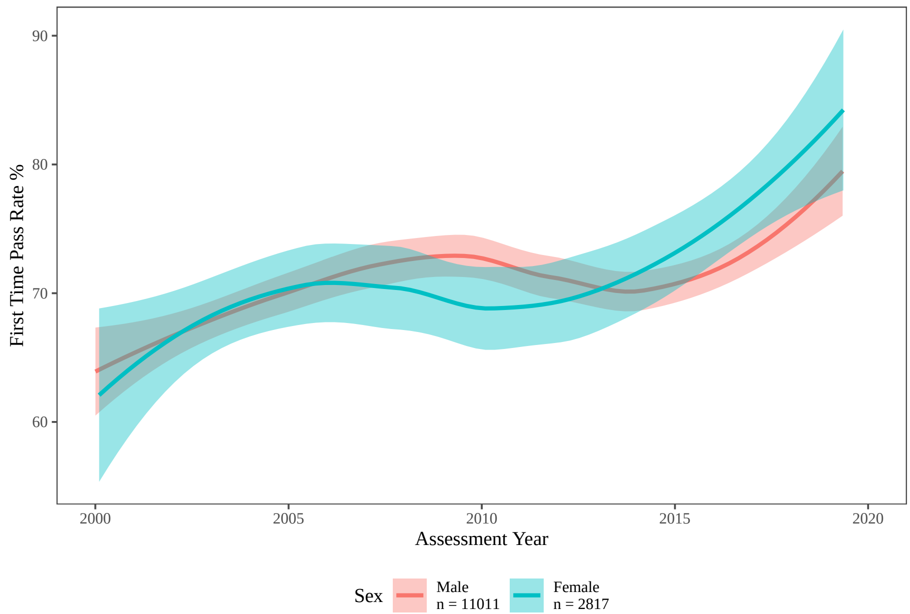

Figure 3.1: Pass rates for female and male candidates assessed since 2000 (n = 13828).

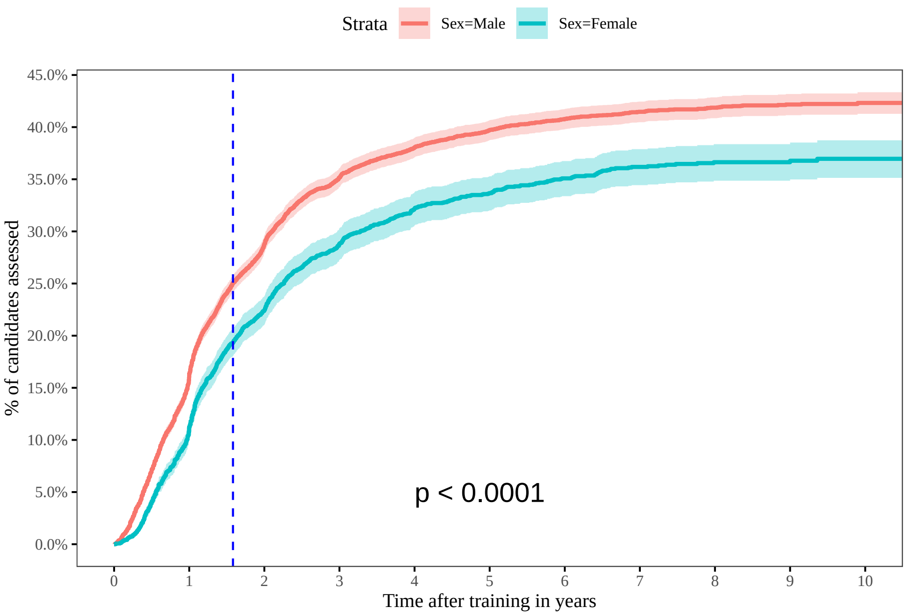

Figure 3.2: Survival rates for female and male candidates post-training.Candidates trained 2009-2019. (n = 15635) Blue dashed line at mean time to assessment.

3.1.2 Pathways

Need to add something about talent ID and development

- Some studies focus on understanding the exact relationship between a small number of variables and an outcome, e.g., PhD Delays (van de Schoot et al., 2013)
- Other studies aim to understand which variables, potentially from a large list, are the most important predictors of an outcome.

Having a combination of factors that influence an outcome is not unique to becoming an ML. Recently there have been a number of projects, in the sport domain, that have tried to find the most important factors in the development of athlete in elite pathways. Examples of projects that have sought to identify the most important discriminatory variables for pathways are: The Great British Medallist Project (GBM; Güllich et al., 2019; Hardy et al., 2017; Rees et al., 2016); “Game Changers” Discriminating Features within the Microstructure of Practice and Developmental Histories of Super-Elite Cricketers (Jones et al., 2019; Jones, 2019).

3.1.3 This paper

- We identified 166 variables that were deemed as potentially important to the completion of the ML qualification in a large qualitative study (see results of Chapter 2 **LINK?? APPENDIX**) and a workshop with the Mountain Training UKI council.
- We wanted to collect quantitative data from candidates for these variables
- This manuscript has three studies in it:

  1) Item selection
  2) Item reduction
  3) Model identification

  How will this look for the thesis vs a manuscript? And does it matter?

  - Write as thesis chapter

3.2 Study 3 - Identifying the most important factors

The aim of this project was to identify variables that influence completion of the ML qualification. In the preceding studies, we created a survey tool of reasonable length to collect data for potentially important variables. In this study, we used this survey tool to collect data in order to identify the most important discriminatory variables. In this study we aimed to identify a feature subset for each of the following classification problems:

1. Male candidates who are assessed 18 months after their training from those who are not.
2. Female candidates who are assessed 18 months after their training from those who are not.
3. Candidates who pass their first assessment from those who do not.

This results for study are therefore presented in three sections, one for each of the classification problems above.

3.2.1 Method

3.2.1.1 participants table

3.2.1.2 Participants

A separate sample of ML candidates were used for this study. We contacted all candidates who attended their first ML training course in 2017 or 2018, inviting them to participate in the study, 1030 candidates started the survey and 440 completed the survey (15.35% response rate). Of the 440 responses, 155 were

Table 3.1: Participant descriptive statistics.

Sex	n	Age	white	Assessed	Assessed within 18 months	Passed first time
Female	246	NA	219 (89%)	71 (28.9%)	61 (24.8%)	59 (83.1%)
Male	502	39.4	446 (88.8%)	187 (37.3%)	171 (34.1%)	159 (85%)

Table 3.2: Candidates pathway progress when completing the survey.

Sex	Assessment	18 months post-training	n
Female	post	post	12
		pre	22
	pre	post	19
		pre	102
Male	post	post	36
		pre	55
	pre	post	35
		pre	159

from female candidates (M_{age} = 36.02 ±11 years) and 285 were from male candidates (M_{age} = 40.88 ±12.75 years). These candidates had been trained by 70 different providers and those who had been assessed, had been assessed by 52 different providers.

When responding to the survey, candidates would have been at different stages in the pathway. Some candidates responded to the survey prior to an assessment and other responded having been assessed, some candidates would also have completed their training course at least 18 months before responding to they survey and the remainder would have been responding to the survey within 18 months of their training course (see Table 3.2 for a summary).

In this study we analysed the data for getting to assessment within 18 months of training from female and male candidates separately as the completion rates for each group are different[2]. Each of the analyses had different inclusion criteria and subsequently, used a different subset of candidates. Details of the candidates included in each data set are presented in the sections below.

3.2.1.2.1 Getting to assessment within 18 months of training - Male candidates

[2]In Study 2 we split candidates according to whether or not they had DLOG data, which meant that it was not possible to then split each of these groups by gender as well because some groups would have been too small to analyse.

There were 71 responses from male candidates who completed the survey more than 18 months after their training course (i.e., retrospectively), 6 of whom had been assessed more than 18 months post-training[3], 33 of whom had been assessed within 18 months of their training course and 32 who not been assessed at the time of completing the survey[4]. Therefore, we were able to create a set of *training data* ($n = 54$), which we could use to select variables and a set of *test data* ($n = 10$, with an equal split of candidates who had and had not been assessed).

In addition to this, 164 male candidates completed the survey less than 18 months after their training but at the time of writing were at least 18 months post-training. These candidate formed the *validation data* set.

3.2.1.2.2 Getting to assessment within 18 months of training - Female candidates

The data used for this analysis were collected from 27 candidates who had been assessed 18-months after their training ($M_{\text{age}} = 36.26 \pm 11.05$ years) and 27 who had not ($M_{\text{age}} = 37.98 \pm 11$ years). We received fewer responses from female candidates, therefore we combined the retrospective and prospective data as neither group would be large enough on its own. In each group there were 10 candidates who completed the survey retrospectively (i.e., $>$ 18 months post-training) and 17 who completed the survey prospectively (i.e., 12-18 months post-training).

3.2.1.2.3 Passing first time

The data used for this analysis were collected from 46 candidates, 34 of whom had been assessed prior to completing the survey and 12 of whom had not been assessed before completing the survey. As with the data in female candidates getting to assessment, we combined the retrospective and prospective data to increase the sample size[5]. Twenty three of the 46 candidates passed their assessment first time. Of the 23 who did not pass, 6 completed the survey prospectively. Two of the 23 candidates who did not pass withdrew from their first assessment, none failed, and the remainder were deferred. Seven of those who were deferred only needed to log additional days.

> Make sure the numbers add up - they don't at the moment, I think this is because of the removal of candidates/variables

[3] These candidates were not included in the analyses as the wording of the questions they answered meant that their responses were not comparable to those who had been assessed within 18 months and those who had not been assessed.

[4] Candidates who had not been assessed within 18 months of their training course but had been assessed prior to completing the survey were excluded from the analysis as the wording of the questions shown to them meant they would not be comparable to the other candidates.

[5] We have run the analyses on just the retrospective data, which allowed us to include some variables about candidates' experiences of assessment, but none of these variables were selected in the best discriminatory subsets, nor were the classification rates significantly higher.

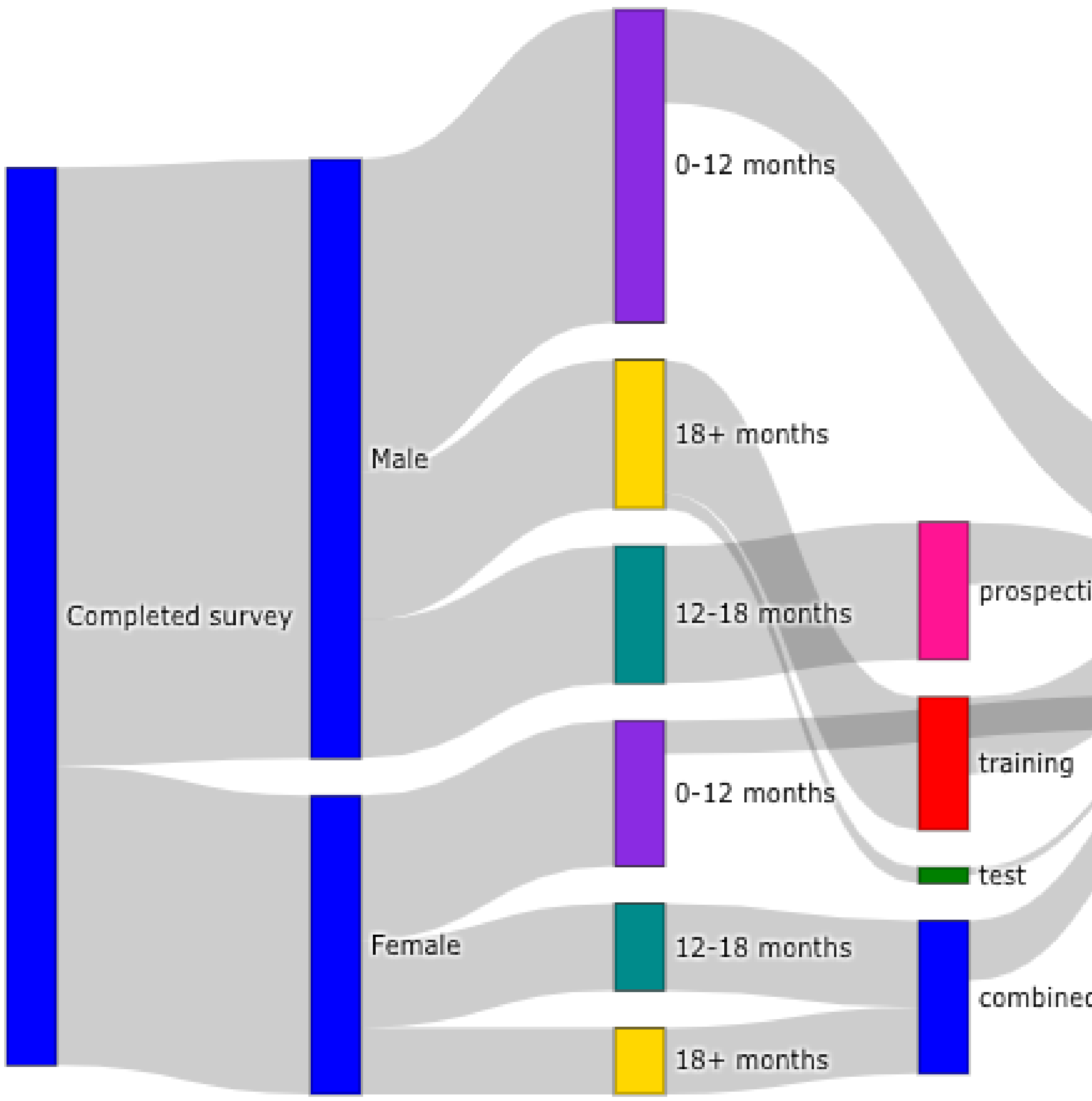

Figure 3.3: Study 3 participants. For simplicity, candidates who have not been assessed have not been added to this figure as a final group, therefore it can be assumed that candidates not progressing from one node to another have not been assessed.

3.2.1.3 Measures

The survey tool developed in Study 2 was used to collect data. See Table 3.7 for a full list of variables included in the survey.

3.2.1.4 Procedure

3.2.1.5 Analytical method

In this study we were interested in identifying the best combination of discriminatory variables and as such, we used a similar but different method to the one used in Study 2. The three steps were the same, but the way in which we created the data sets and some of the technical specifications for the parameters in the algorithms were different.

3.2.1.5.1 Preprocessing

For each classification problem, the ideal way to perform the analyses would be as follows:

1. Given N cases, randomly select x, where $\frac{N}{3} > x > \frac{N}{10}$ cases for each category of the criterion variable to be held-out as a test data set D_{test} and the remaining candidates become the training data D_{train}.
2. Prepare both the D_{train} and D_{test} data sets separately (e.g., standardising the data).
3. Perform the feature selection process using D_{train}.
4. Carry out the classification process using D_{train} using k-fold cross validation.
5. Carry out the classification process on D_{test} using D_{train} to train the classification model.

Using the same data to train and test a model leads to the risk of over-fitting and classification rates being artificially inflated as all of the data has been "seen" during the feature selection stage. This phenomenon is know as "peeking" (Kuncheva and Rodríguez, 2018; Reunanen, 2003; Smialowski et al., 2010) and can be mitigated by holding some data out of the feature selection stage (i.e., D_{test}).

3.2.1.5.2 Feature selection

Feature section in this study was carried out in a similar fashion to that in Study 2, however we made three changes. The first and most substantive change was the way that decentralisation for feature selection was carried out. For this study, features were vertically distributed (i.e., split by features and not by case; Bolón-Canedo et al., 2015b). For each classification problem we created seven disjoint feature subsets (for a list of survey variables in each subset see Table 3.7). The second change was the use of the Fast Based Correlation Filter (FCBF; Yu and Liu, 2003) instead of CFS; FCBF is a more conservative subset evaluator than CFS. A final minor difference, is that we used the $k = 6$ for the Relief-f algorithm.

We then carried out the feature section process as described in Study 2 on each of the feature subsets above as well as the full set of features, resulting in at least two feature subsets, one for variables selected by at least two FS algorithms (2s) and one for variables selected by at least three FS algorithms (3s), for each of the feature (sub-)sets. Where there were at least five variables selected by all four FS algorithms we also retained that feature subset (4s). We then created three sets of omnibus feature subsets for each classification problem where each of the 2s, 3s, and 4s were combined to form a previously unseen feature subset; an example of the omnibus features subsets for getting to assessment is presented in @ref(??).

3.2.1.5.3 Initial classification

We performed a classification experiment for each of the classification problems using WEKA's Experiment Environment (Bouckaert et al., 2018; Frank et al., 2016). For each feature subset identified in the FS stage the training data were classified using LOO-CV and the same four classification algorithms as in Study 2 across 10 runs. This process returned a mean classification rate for each feature subset and classifier.

3.2.1.5.4 Final classification

3.2.1.5.5 Model selection

The feature selection process yielded 28 getting to assessment models for both male and female candidates and 26 first time pass models. For each of the classification problems listed above, we selected the "best" models. It is important to recognise that these three solutions are not the only useful ones, however, they were the models that best classified the training data. It is also important to note that we considered the classification profile for each model, rather than just the mean score. It is not uncommon for one classifier (often J48) to perform much worse than the others, therefore if a model performed well with three classifiers and poorly with another that model was preferred to one that performed better on average (i.e., had a greater mean classification accuracy). For example, consider the classification profiles of the following models: Model A - NB = 85, SMO = 90, IBk = 85, J48 = 50 (mean = 77.5) and Model B - NB = 80, SMO = 80, IBk = 80, J48 = 80 (mean = 80). In this example we would prefer Model A to Model B.

3.2.1.5.6 Model testing

For the test data set, we used the training data to train the model to predict the class for the test data. The performance of each classifier within the ensemble was then evaluated by the classification rate (percentage accuracy) and inspection of the confusion matrix.

Table 3.3: Group 5 male candidates getting to assessment within 18 months of training, unstandardised group descriptive statistics

Variable
Perceived available time to become an ML
Resilience
Importance of becoming an ML
Relative importance of becoming an ML
Progress towards becoming an ML
Relative progress towards becoming an ML
Relative efficacy of becoming an ML
Recalled understanding of the qualification pre-training
Expected time (in months) to assessment at start of training
Expected time (in months) to assessment at the end of the training course
Intention to complete at the end of training
Experience of social change post-training
Actual received availability of emotional support (in the last week/week before assessment)
Have done nothing to prepare
Perceived preparation in the last six months/six-months before assessment
Pre-assessment efficacy to "plan a mountain day that is appropriate for the group"
Difference between ideal and pre-assessment efficacy to "act according to my responsibilities to other

For the validation data set, we assigned a *voted predicted class* for each candidate based on the average predicted class across the classifier ensemble[6] for each model developed with the training data. We used the classification rate based on the voted predicted class to evaluate the performance of the model.

3.2.2 Results

3.2.2.1 Getting to assessment within 18 months of training

3.2.2.1.1 Male candidates

3.2.2.1.1.1 Model selection

We tested 28 models using the training data; the performance of each model can be seen in Table 3.9. There were three models tied with the best performance, containing 17 unique features. No model contained all 17 features, however, 0 features were included in all three models and 6 features were unique to just one of the three models. Figure 3.4 shows the normalised group means for the training data of the 17 features selected across the models and Table 3.3 shows the unstandardised descriptive statistics.

[6]We also used the Multilayer Perceptron (MLP; Bishop (2006)) classifier for these classification analyses to ensure that there were no ties amongst the predicted classes for a given object.

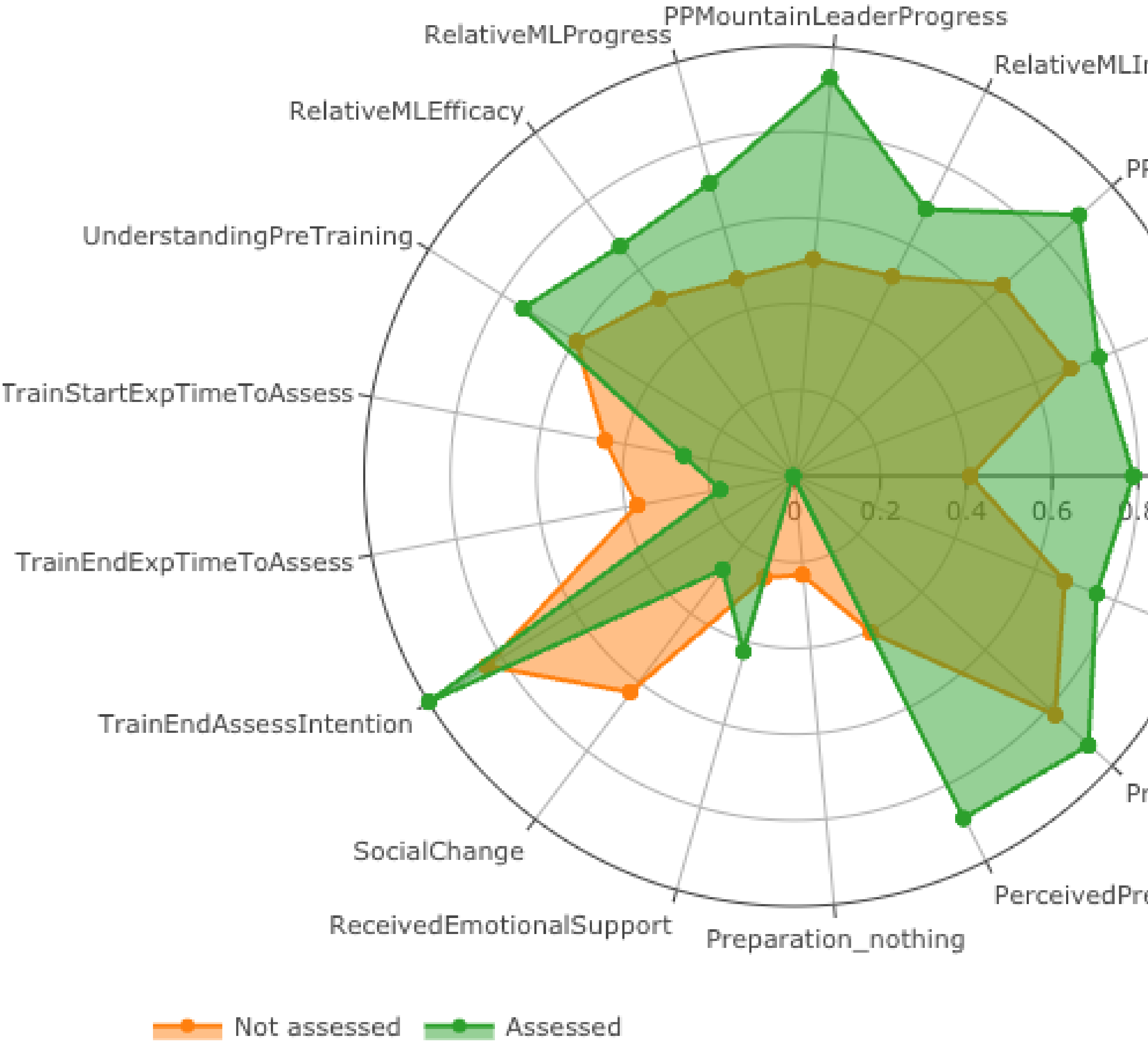

Figure 3.4: Normalised training group means for male candidates getting to assessment within 18 months of their training course.

Table 3.4: Group 5 male candidates getting to assessment within 18 months of training, prediction model performance (n = 164).

model	NB	SMO	IB6	J48	MLP
omnibus_survey_3s_3s	70.73	72.56	69.51	73.17	73.78
omnibus_2s_3s	70.73	73.17	71.34	73.17	73.78
all_in_3s	70.73	73.17	72.56	73.17	73.78

Table 3.5: Group 5 male candidates getting to assessment within 18 months of training, prediction model performance

survey_time	assessed_18m	assessed	all_in_3s	omni_2s_3s	omni_survey_3s_3s	n
pros	FALSE	FALSE	74.67	74.67	74.67	75
pros	FALSE	TRUE	26.67	26.67	26.67	15
pros	TRUE	TRUE	63.16	63.16	63.16	19
retro	TRUE	TRUE	90.91	87.27	87.27	55

3.2.2.1.1.2 Model testing

We tested the three models selected above on the previously unseen test data. Across all five classification algorithms, each model classified the test data with 90% accuracy. In each of the three models, case 8 was misclassified by NB, SMO, and IBk and case 3 was misclassified by J48. The fact the models made the same errors is unsurprising given the number of shared attributes. The performance of these three models on the test data is evidence that the model is not over-fitted to the training data, thus increasing our confidence that the variables selected are important discriminatory variables.

3.2.2.1.1.3 Model validation

In these data, the performance of the classification algorithms was consistent across both the classifiers and models, but was lower than in the test data (Table 3.4).

Table 3.5 shows that the models are good at classifying candidates who have not been assessed (CR $\approx$ 75%) and those who have been assessed within 18 months and completed the survey after having been assessed (CR $\approx$ 90%), but is poor at classifying those who are assessed after more than 18 months (CR $\approx$ 27%) and those who are assessed within 18 months but completed the survey prior to being assessed (CR $\approx$ 63%).

3.2.2.1.2 Female candidates

Table 3.10 shows the classification rates for all models

DLOG t6 and DLOG t12 seem to the same

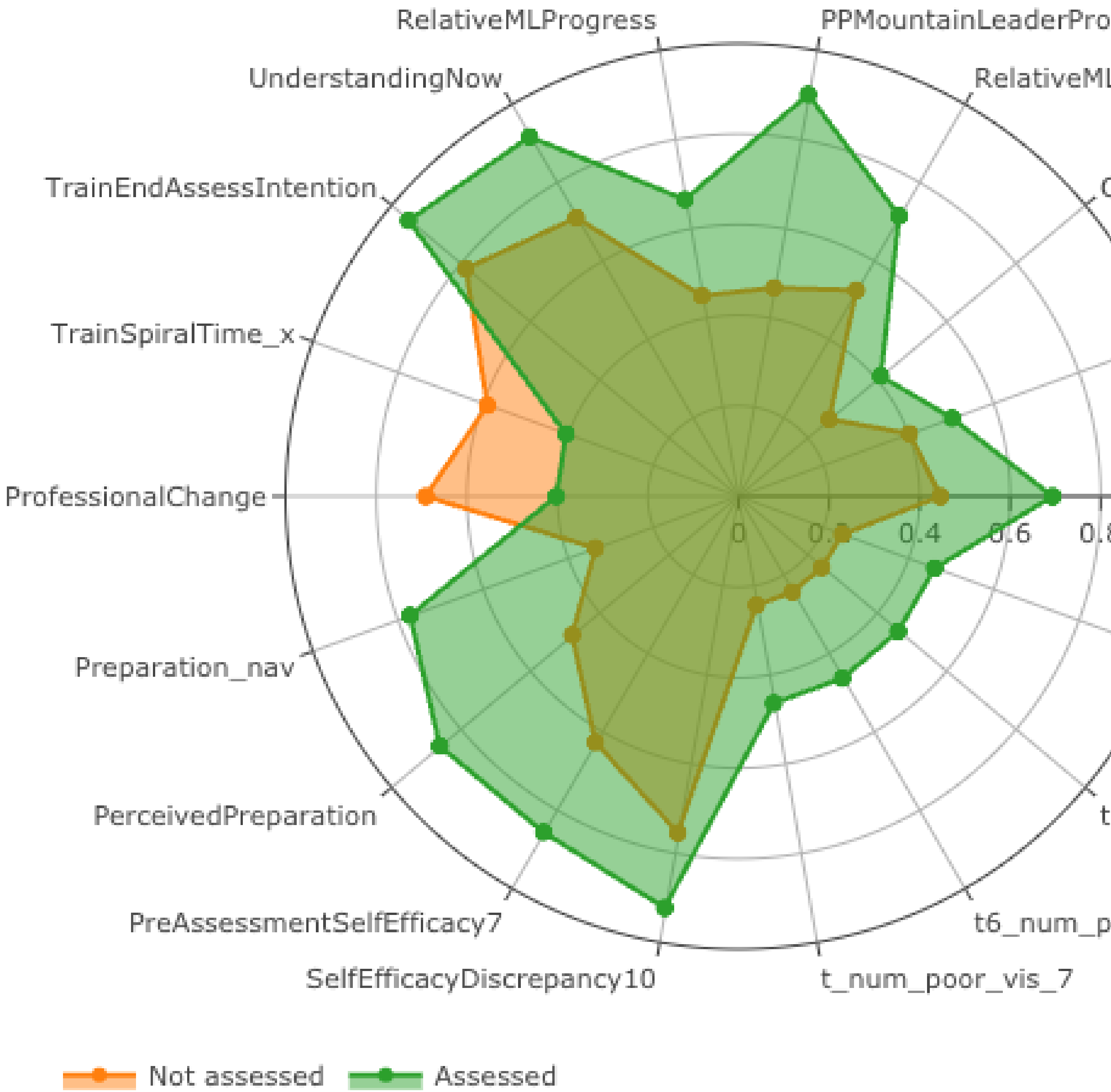

Figure 3.5: Normalised training group means for female candidates getting to assessment within 18 months of their training course.

Table 3.6: Group 5 female candidates getting to assessment within 18 months of training, unstandardised group descriptive statistics

Variable
Perceived available time to become an ML
Number of QMDs logged by ideal self before assessment
Number of QMDs logged by ought self before assessment
Relative importance of becoming an ML
Progress towards becoming an ML
Relative progress towards becoming an ML
Understanding of the qualification when completing the survey
Intention to complete at the end of training
TrainSpiralTime_x
Experience of professional change post-training
Preparation included navigation
Perceived preparation in the last six months/six-months before assessment
Pre-assessment efficacy to ”navigate to a chosen point on a map in any conditions, night or day”
Difference between ideal and pre-assessment efficacy to ”act according to my responsibilities to other
t_num_poor_vis_7
t6_num_poor_vis_7
t12_num_poor_vis_7
t18_num_poor_vis_7

3.2.2.1.3 Passing first time

- Only top 10 features for previous courses & omnibus_DLOG_3s

Table 3.11 shows the classification rates for all models

3.2.3 Discussion

The studies presented in this report aimed to identify important factors that discriminated candidates who (a) having been trained, went on to be assessed within 18 months of training from those who did not, and (b) having got to their first assessment, pass first time from those who do not. To achieve these aims we considered a wide range of potentially relevant variables. The results presented show that there is no one single factor that is important for discriminating candidates and in fact there are some important commonalities between groups, which are likely fundamental for the successful completion of the ML qualification. Some of the discriminatory variables are common to both stages of completion, or to both female and male candidates getting to assessment.

3.2.3.1 Model performance

3.2.3.1.1 Male candidates - Getting to assessment

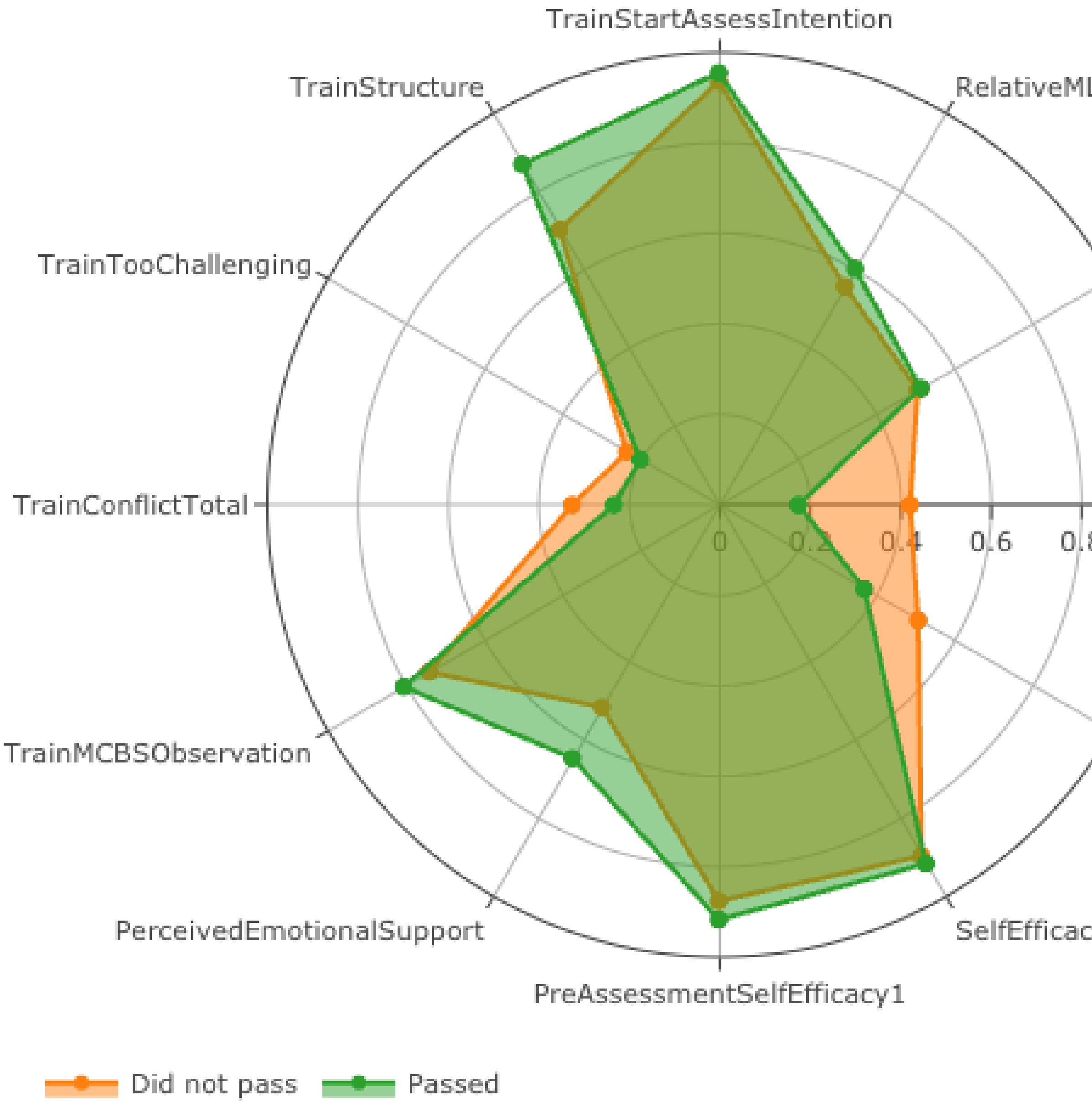

Figure 3.6: Normalised training group means for candidates passing their first assessment - survey variables.

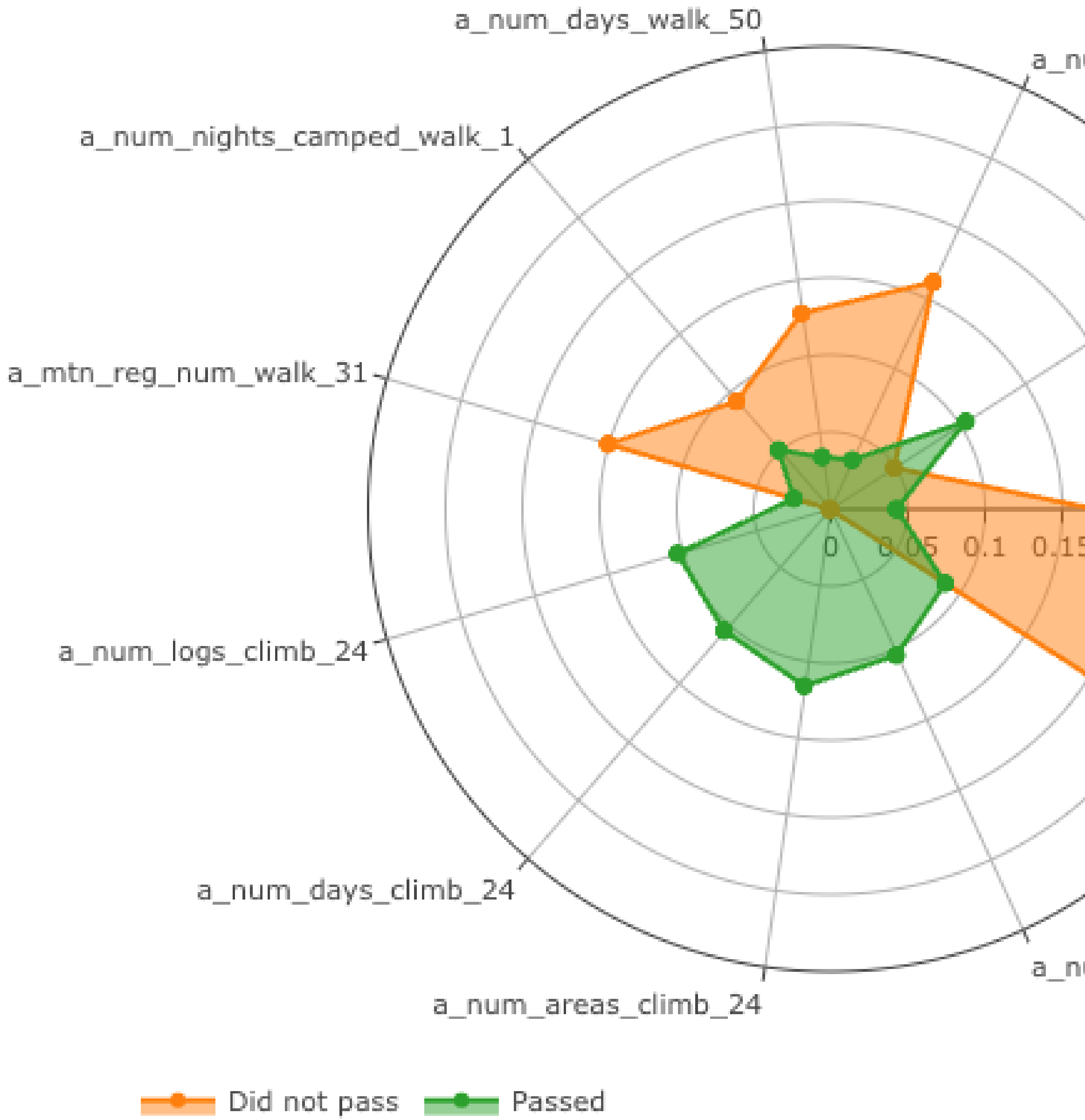

Figure 3.7: Normalised training group means for candidates passing their first assessment - DLOG variables.

The classification rates for both the training and test data were very good (i.e., approx 90%), however the predictive classification rates were lower (approx 70%). This suggests that the model developed is not over-fitted to the training data, as there is little difference in the training and test classification rates.

There are three main reasons this may be: firstly, in the training and test data candidates who had been assessed more than 18 months post-training were excluded from the analyses, these candidates would most likely be misclassified as assessed within 18 months of training **EXPLAIN**. Secondly, these data were collected from candidates who had completed their training less than 18 months prior to training

there are (at least) two plausible reasons for the lower classification rate in the predictive data. Firstly, it is possible that candidates are creating narratives in their minds based on whether or not they have been assessed and that had they answered the survey before being assessed, then their answers may have been different. A second explanation is that because candidates were not at the same point in time when they answered the survey, their answers were different to those they would have given six-months later. For example, a candidate may not have felt that they had made much progress towards becoming an ML 6-12 months post-training and then may make progress in the 12-18 month period.

This is not surprising given that group was not included in the training data, thus is arguably a different classification problem. Analysing the data collected from candidates who had not been assessed when the completed the survey but were trained over 18 months ago, may identify a different model. Ideally data would be collected in a proper prospective fashion.

The results from this dataset do allow us to place more confidence in the models selected and many of the classification errors may not be important to Mountain Training as some of the candidates who are being misclassified are still going on to be assessed (most within 24 months of their training).

3.2.3.2 Features selected

3.2.3.2.1 GTA Commonalities

Several features were common to both the male and female getting to assessment models. At least one model selected for each GTA classification problem contained the following features:

- Perceived available time
- Relative ML Importance
- ML Progress & Relative ML progress
- Train end assess intention
- Social/professional change
- Perceived preparation (& prep_nothing for males)
- SelfEfficacyDiscrep10 - act according to responsibilities to others

The results of the pattern recognition analyses suggest that how becoming an ML fits into candidates' lives is important in discriminating both female and male candidates who are assessed within 18 months of their training course. Candidates who are assessed within 18 months of their training course were more likely to have: felt that they had enough time to become an ML, felt that becoming an ML was more important than other life goals when registering (becoming an ML was also more important for male candidates who were assessed within 18 months than for male candidates who were not), have made more progress towards becoming an ML, have made more progress towards becoming an ML than towards their other life goals, had a stronger intention of being assessed at the end of their training course, have experiened less change in their life post-training (profesisonal for females and social for males), felt that they had done more to prepare effectively for an assessment.

- If a candidate feels that becoming an ML is important to them, or that it is more important than other life goals, they may be more likely to commit resources towards it.
- Time is one such resource, thus those candidates who feel that beomcing an ML is important to them may be more likley to feel that they have more available time to do so.
- Having more time would also allow candidates to spend mroe time preparing and therefore feel like they had done more to prepare effectivley for an assessment.
- Whilst we did not collect any data on the exact nautre of changes that candidates' experienced post-training, the fact they are associated with them not having been assessed would suggest that these changes did not allow candidates to spend more time preparing to become an ML. Candidates who experience changes in their lives post-training

3.2.3.2.2 Male GTA

3.2.3.2.3 Female GTA

3.2.3.2.4 FTP

The results presented in Getting to assessment: Male candidates suggest that how becoming an ML fits into male candidates' lives is important when considering the likelihood of them being assessed. If a candidate **feels that becoming an ML is an important life goal, generally or relative to other life goals**, they may be more likely to commit time and resources towards it, thus may feel that they can prepare for an assessment in a shorter period of time, which for many, would include revisiting more technical areas of the syllabus like river crossings or practising skills they rarely use like emergency rope work. Candidates who felt that they had more **available time** to become an ML, had done more to **effectively prepare for an ML assessment**, had made more **progress towards becoming an ML**, and were **more confident that they**

could become an ML than to achieve other life goals were more likely to have been assessed 18 months after their training course.

Some candidates are less certain in their **understanding of the purpose of the ML qualification prior to their training course** and may be attending in order to find out more about the qualification, whereas those who are more certain of the purpose are more likely to be doing it in order to progress to an assessment. The **strength of candidates' intentions to be assessed at the end of their training course** being an important discriminatory variable is in line with the *Theory of Planned Behaviour* (Ajzen, 1991). The Theory of Planned Behaviour suggests that intentions are the strongest predictors of behaviour and that the strength of these intentions also predicts the behaviour (Armitage and Conner, 2001).

The strength of a candidate's intention to be assessed at the end of the training course may be more important than their intention at the start because the candidates who were less sure of the purpose of the ML qualification would have had less information to base their intention on. This position is supported by the fact the correlation between being assessed 18 months post-training and the intention to be assessed at the start of the training course (0.1709069) is lower than the correlation between being assessed 18 months post-training and the intention to be assessed at the end of the training course (0.3651245). The results in Expectations and intentions support this, including using prospective data and retrospective data from female candidates, which suggests the strength of intention is important for all candidates, despite not being one of the most important discriminatory variables for female candidates.

Candidates who **expect it to take them longer to get from training to assessment** are less likely to be assessed within a given period[7]. Candidates may also expect it to take them longer as they either have less available time, live further from the mountains, or a combination of the two, making it more difficult to fit into their lives. If candidates who expect to take longer do take longer, then there will be more opportunities for things to get in the way of them pursuing that goal and becoming barriers to completion.

Further, experiencing **social change** after a training course may mean that candidates have more or less available time, or have changes in their priorities. The question used in the survey did not ask if candidates had more or less resources (e.g., available time) because of this change, however given that more social change a candidate experienced, the less likely they were to be assessed within 18 months, it would be reasonable to assume that these social changes are more likely to leave candidates with less, rather than more, resources to become MLs.

- PhD delays study found that the effect of having children was greater for men than women (van de Schoot et al., 2013), supporting another that suggests the same for *** (Waite, 1995).

In our analyses we used the time of year that courses took place as a proxy measurement of weather and daylight hours. We would expect courses near

[7]Of the 83 candidates not assessed within 18 months of their training, only 27 expected it to take them more than 18 months.

the New Year to have worse weather and less daylight than those nearer to the middle of the year. Given that candidates who were **trained closer to the middle of the year (i.e., June/July)** were more likely to have been assessed 18 months after their training course, it is likely that better weather and more daylight on the training course provides candidates with a more positive experience and possibly a better learning environment. To investigate this further, weather data (held on CMS) and daylight hours data should be included in the feature selection stage of a reanalysis of these data.

There is a broad literature reporting the benefits of resilience (e.g., Seery and Quinton, 2016). Becoming an ML is a difficult process which requires the investment of time, energy, and money and most candidates will have to deal with setbacks during this process. Candidates who are more **resilient** will be better able to overcome adversity (Smith et al., 2008). For example, bad weather on a training course or changes in life circumstances that become barriers to becoming an ML). It is also a central tenet of *self-efficacy* theory that people with firmly established self-efficacy beliefs are more resilient (Bandura, 1997) as the stronger self-efficacy beliefs are, the easier they are to maintain following disconfirming events.

One would normally expect the availability of social support to be a positive influence, however the results in this study suggest that having higher levels of ***perceived esteem support*** means that candidates are less likely to have been assessed 18 months after their training course. One explanation for this is that candidates who do not feel that they need esteem support answer this question in a different way to those who do (i.e., they don't perceive it as available), therefore those who feel they need it score more highly and with less variation in their responses. Another explanation is that esteem support may be reminding male candidates who are unprepared that they are not ready for an assessment. Similar findings, involving the somewhat paradoxical effects of psychological skills and strategies on performance (Roberts et al., 2013), have been reported elsewhere in the literature and do highlight that some support strategies might need to be utilised with caution.

Candidates who **feel less able to look after themselves and others than they would in an ideal world on steep ground and crossing rivers** may feel that they are not ready to pass an assessment and therefore not attend one. For a number of candidates, these skills will be the most specialist mountaineering skills they possess and will have little reason, beyond passing an ML assessment, to practice them. Unless these candidates have spent time deliberately preparing for an assessment, it is likely that they will feel less confident than they would like to at assessment, that they can successfully demonstrate these skills.

3.2.3.3 Female candidates - Getting to assessment

As with the results for male candidates, **how important becoming an ML is to a candidate, relative to other life goals**, is an important discriminatory variable for female candidates. We would expect this variable to have the same implications as those already discussed for male candidates. Again, the more **progress that candidates have made towards becoming an ML**, the more likely they are to feel that they have **prepared effectively for an ML assessment** and in doing so, they will have gained experience that boosts their confidence in their abilities to perform task related to the assessment. It is likely that **professional change** will have similar effects for female candidates as social change does for male candidates.

We asked candidates to give two reasons that they had registered for the ML qualification. For their first reason, most candidates said that they had registered in order to become an ML (n.b., this is an extrinsic participatory motive). The candidates who gave an ***extrinsic participatory motive*** **for their second motive** (e.g., "Qualify as an ML)") rather than a more intrinsic one (e.g., to spend more time in the mountains) were more likely to have been assessed 18 months after their training course. This finding suggests that having more than one extrinsic participatory motive is important for candidates getting to assessment.

Goal setting has been shown to improve outcomes in a number of domains (see Weinberg and Gould, 2014, p 356). One way that **goal setting facilitated by training course staff** may have helped candidates is by enabling them to maximise the benefits of the time that they spent consolidating their skills and preparing for an ML assessment after the training course. In addition to this, goal setting may have made it more likely that candidates would prepare for an assessment. The more specific these goals are, the more they will have focused candidates' attention and efforts towards being at the right level to pass an assessment. Further, *goal setting will have helped facilitate mastery experiences (i.e., having an experience where one is successful)*, the strongest source of self-efficacy (Bandura, 1982); thus, this goal setting will have helped female candidates develop their confidence, which as discussed below, is key for female candidates getting to assessment.

If candidates feel that becoming an ML is important to them, they may also feel that it is important that they are good enough to pass when they get there. This suggestion helps to explain why candidates who were assessed felt that **ideally, they would have a higher number of QMDs at assessment**. Another explanation could be that candidates who have not received goal setting support have fewer clear goals and do not feel that they can use the time as efficiently, therefore feel that they would ideally have more QMDs before being assessed.

The results presented in Female candidates: Getting to assessment and ML related self-efficacy show that female candidates who are assessed within 18 months of their training have higher levels of **self-efficacy pre-assessment** than those who are not and that these higher levels of self-efficacy are associated with experience gained after the training course. These items are about areas of the syllabus relating to hazards and emergency procedures, where mistakes may have serious and immediate consequences for other people. It may be especially important for course staff to help female candidates set goals that help them develop their confidence to perform these tasks.

Discrepancies between the ideal and post-training levels of self-efficacy were not selected as important discriminatory variables, whilst three of the **pre-assessment self-efficacy** items were. This would suggest that it is not the discrepancy that is important, but the pre-assessment levels of self-efficacy, which will be influenced by candidates' experiences and how much preparation they feel that they have done. This hypothesis is supported in Sex differences where there is evidence of a positive relationship between experience and confidence, which is stronger for female candidates than it is for male candidates.

It is both interesting and important to note, that 10 of the 11 the features in this discriminatory subset relate to the consolidation period. Considering this combination of variables, the timing of them, and the relationship between the number of QMDs and pre-assessment self-efficacy, the importance of female candidates gaining additional and relevant experience after their training course becomes paramount.

3.2.3.4 Passing first time

The **further candidates live from a mountainous region**, the more difficult it will be for them to gain relevant experience. Furthermore, it is also less likely that they will be able to access support specific to becoming an ML as it is less likely that becoming an ML is normal in their social context.

It is clear from analyses not reported here that the first time pass rate for the ML qualification is lower for **non-White-European** candidates than it is for White-European candidates[8] and also that the proportion of non-White-European candidates who are assessed is *much lower* than the proportion of White-European candidates who are assessed[9]. There are many plausible explanations for this, which may include social, cultural, and economic factors. However, there is little empirical evidence to support any of them at the moment and it is beyond the scope of this report to examine this issue further.

The facilitation of goal setting by course staff was also an important factor for passing first time. In addition to helping candidates set goals, the ***provision of structure*** **by training staff**, by making it clear to candidates what they need to do to be pass an assessment, was important. The provision of structure may have benefited candidates by helping them to set *very clear and specific goals*, which are more effective than broad and/or vague goals for influencing behaviour change (Gould, 2005).

There are a number of reasons that **extraversion** may be linked with passing, including differences in levels of physiological arousal, which can influence the breadth of perceptual cues that individuals pay attention to, and decision making (Hardy et al., 1996). Extraversion has also been linked with effective leadership (Judge et al., 2002). There is also evidence that *goal setting reduces the distractibility of extraverts*, helping them maintain focus in training (Woodman et al., 2010), therefore, goal setting may be particularly important for extraverted candidates.

The ML assessment is a very stressful experience for many candidates. It is unsurprising that ***received emotional support*** and **perceived esteem support available** are positive predictors of passing. Having these types of social support may help candidates cope with the pressure of assessment (Freeman et al.,

[8]Analysis of data on CMS shows that the pass rate for non-White-European candidates has been lower than for White-European candidates since at least 2010.

[9]This is in general and not just after 18-months.

2014, 2011). However, as seen above, perceived esteem support is a predictor of male candidates not getting to assessment. These findings would suggest that esteem support should be used sparingly, or only in the right context (i.e., when candidates are ready to be assessed).

Seven of the 23 candidates who did not pass their first assessment were *only* deferred because they had too few **Quality Mountain Days in their logbook at assessment**. *It is important to highlight that the features presented here discriminate between candidates who do and do not pass their assessment, not between candidates who are and are not good enough to pass an ML assessment, in terms of their skills and decision making.* If we removed these particular candidates from the sample, we would have too few cases to perform the analysis, therefore, it is difficult at this juncture to answer the question “Is having more than the minimum experience beneficial for passing an ML assessment.” If anything, it is evidence that one can pass the practical element an ML assessment with *fewer than* 40 QMDs.

The results presented in Passing first time also suggest that candidates who include **Quality Hill/Moorland Days** in their DLOG are less likely to pass. Whilst **it is unlikely that this experience is detrimental to their performance at assessment**, Quality Hill/Moorland Days are not as relevant as QMD experience. One explanation for this finding is that candidates who feel they have a weak logbook want to show all the experience that they feel is relevant, whereas a candidate who feels they have a strong logbook may only feel the need to include the experience they feel is most relevant. Further, candidates who live further from the mountains may be trying to prepare for an ML assessment in non-mountainous terrain as it is more accessible to them.

Nine of the 10 candidates who **attended a Mountain Skills** course prior to being assessed and responded to the survey passed their first ML assessment[10]. This suggests that additional structured training helps candidates to successfully prepare for an assessment.

When considering the discriminatory features presented above in a holistic manner, it is important that whilst preparing for their assessment, candidates gain enough relevant experience in the consolidation period, using clear and specific goals developed from training. In addition, it is vital that they are able to cope with the pressures of the assessment process, drawing not only on their experience relevant to the ML qualification (i.e., QMDs), but also social support when necessary.

3.2.3.5 Limitations

Several limitations can be identified in this project. Firstly, most of the data used was collected retrospectively. Retrospective data will be less accurate as time increases between the event and when participants are sampled, and

[10] The candidate who did not pass attended a Mountain Skills course 35 days before the start of their assessment and their training course 107 days before their assessment (all with the same provider). They also had an additional seven days experience (Dartmoor & Snowdonia) between the Mountain Skills course and their assessment.

people may create their own narrative retrospectively which may or may not reflect reality. An example of this could be a candidate who did not pass their first assessment attributing their failure to the coaching (or lack thereof) they received on their training course.

Secondly, there is some evidence of sampling bias in the data used to identify the important discriminatory factors for both getting to assessment and passing. The proportion of female and male candidates who did get to assessment within 18 months of their training course is not the same in the retrospective data (females = 23.21% and males = 41.35%) as it is in the population of candidates trained in the same period (females = 19.02% and males = 30.22%). In addition to this, the proportion of males who did not pass their first assessment is not the same in the retrospective data (13.5%) as it is in the prospective data (19.6%) or in the population[11] (19.8%); there is no evidence of the same problem in the data collected from female candidates. The simplest explanation for this is that candidates who are not assessed and male candidates who do not pass their first assessment are less likely to ***retrospectively*** respond to the survey.

Whilst there may be a subset of candidates that are not represented in the data collected as part of this project, we believe that the findings presented in this report can be used to make a positive impact on the completion rate of the ML qualification. This belief is based not only on the analyses of retrospective and prospective data presented here, but their congruence with the results from the initial qualitative study and existing literature.

Reanalysis of these data in the future should mitigate this sampling bias so that the response rate in the prospective data is similar to that in the population and reduce the impact of recall bias. However, a truly prospective study that collected data from candidates at registration, training, and during their consolidation would likely overcome the limitations described above.

- Indicators not measures
- Retrospective
- Dichotomising continuous data?

3.3 Discussion

3.4 Future

- Ethnicity
- Self-efficacy
- Life changes

[11] Candidates who were first trained after 2016.

3.5 Conclusion

3.6 Study 2 supplementary information

3.6.1 Variables

- Add column for which groups were asked each variable (including Group 5)

3.6.2 Study 2 classificaiton rates

3.7 Study 3 supplementary information

3.7.1 Example ominbus feature subsets

1. Omnibus survey

   a) Survey 2s = Psychosocial 2s + Training 2s + Consolidation 2s
   b) Survey 3s = Psychosocial 3s + Training 3s + Consolidation 3s
   c) Survey 4s = Psychosocial 4s + Training 4s + Consolidation 4s

2. Omnibus DLOG

   a) DLOG 2s = DLOG t 2s + DLOG t6 2s + DLOG t12 2s + DLOG t18 2s
   b) DLOG 3s = DLOG t 3s + DLOG t6 3s + DLOG t12 3s + DLOG t18 3s
   c) DLOG 4s = DLOG t 4s + DLOG t6 4s + DLOG t12 4s + DLOG t18 4s

3. Omnibus survey and DLOG

   a) Survey and DLOG 2s = Psychosocial 2s + Training 2s + Consolidation 2s + DLOG t 2s + DLOG t6 2s + DLOG t12 2s + DLOG t18 2s
   b) Survey and DLOG 3s = Psychosocial 3s + Training 3s + Consolidation 3s + DLOG t 3s + DLOG t6 3s + DLOG t12 3s + DLOG t18 3s
   c) Survey and DLOG 4s = Psychosocial 4s + Training 4s + Consolidation 4s + DLOG t 4s + DLOG t6 4s + DLOG t12 4s + DLOG t18 4s

Table 3.7: Survey variables

var_description
Perfectionistic strivings
Perfectionistic concerns
Resilience
Robustness
Robustness of confidence
Extraversion
Agreeableness
Conscientiousness
Emotional stability
Openness to experience
Highest education level at registration
Income level
Intention to complete at registration
Expected time (in months) to completion at registration
Employment status at registration
Holiday entitlement (paid) at registration
Perceived travel time (in hours) to the nearest mountainous region at registration
Perceived ease access to the mountains at registration
Aspirations at registration
Registered for multiple qualifications at registration for ML
Participatory motives for registering
Level of intrinsic motivation for registering
Level of integrated motivation for registering
Level of identified motivation for registering
Level of introjected motivation for registering
Level of external motivation for registering
Importance of becoming an ML
Progress towards becoming an ML
Efficacy to become an ML
Relative importance of becoming an ML
Relative progress towards becoming an ML
Relative efficacy of becoming an ML
Perceived available time to become an ML
Understanding of the qualification when completing the survey
Recalled understanding of the qualification pre-training
Intention to complete at the start of training
Expected time (in months) to assessment at start of training
Perception of training course staffs' "observation" skills
Perception of training course staffs' "effective questioning" skills
Perception of training course staffs' "goal setting" skills
Perception of training course staffs' "developmental feedback" skills
Perception of training course staffs' "motivational feedback" skills
Perceived provision of autonomy by training staff
Perceived provision of structure by training staff
Perceived provision involvement of training staff
Perception that the training course felt like an assessment
Type of post-training debrief
Expected change in Expected time (in months) to assessment post-debrief
Perceived understanding of how to prepare effectively for an assessment
Post-training efficacy to "wild camp for two nights in any weather"
Post-training efficacy to "choose appropriate routes whilst leading others in the mountains"
Post-training efficacy to "choose appropriate equipment for mountain walking and explain the choice

Table 3.8: Classification rates for the feature subset with the highest classification rates for each data set (percentage accuracy).

				Classifier			
Analysis	Group	Data	Subset	J48	IBk	NB	SMO
Getting to assessment within 18 months of training	1	With DLOG	3s	74.49	60.20	70.41	73.47
		No DLOG	3s RFE	79.17	76.39	77.78	76.39
	2	With DLOG	4s	75.51	76.53	82.65	78.57
		No DLOG	2s RFE	79.49	79.49	80.77	87.18
	3	With DLOG	2s	66.67	61.90	65.87	68.25
		No DLOG	2s RFE	54.55	66.67	74.24	80.30
	4	With DLOG	3s	71.19	70.34	78.81	77.97
		No DLOG	3s	65.38	71.79	69.23	73.08
Passing first time	1	With DLOG	2s RFE	61.54	67.31	76.92	75.00
		No DLOG	3s	63.89	63.89	61.11	75.00
	2	With DLOG	2s RFE	81.25	81.25	84.38	87.50
		No DLOG	3s	50.00	73.81	64.29	57.14
	3	With DLOG	3s	69.44	72.22	77.78	83.33
		No DLOG	2s	72.50	45.00	62.50	67.50
	4	With DLOG	3s	69.44	72.22	77.78	83.33
		No DLOG	2s	72.50	45.00	62.50	67.50

Note: J48 = J48 Decision Tree, IBk = Instance Based Learning (k=1), NB = Naïve Bayes, SMO = Sequential Minimal Optimization

Table 3.9: Group 5 male candidates getting to assessment within 18 months of training, training model performance

Dataset	n	NB	SMO	IBk	J48	Rating
Psychosocial_2s	21	79.63	83.33	87.04	87.04	Very Good
Psychosocial_3s	11	83.33	85.19	87.04	90.74	Very Good
Training_2s	19	77.78	69.63	57.41	62.96	Modest
Training_3s	9	75.93	77.78	70.37	81.48	Good
Consolidation_2s	21	74.07	77.78	77.78	85.19	Good
Consolidation_3s	10	77.78	79.63	74.07	75.93	Good
omnibus_survey_2s_2s	17	87.04	90.74	88.89	87.04	Very Good
omnibus_survey_2s_3s	11	90.74	94.44	87.04	87.04	Excellent
omnibus_survey_3s_2s	22	88.89	90.74	88.89	87.04	Very Good
omnibus_survey_3s_3s	16	90.74	92.59	90.74	87.04	Excellent
DLOG_t_2s	16	53.70	65.00	50.00	33.33	Poor
DLOG_t_3s	7	55.56	64.81	57.41	59.26	Poor
DLOG_t6_2s	17	53.70	59.44	51.85	42.59	Poor
DLOG_t6_3s	7	64.81	64.81	57.41	57.41	Poor
DLOG_t12_2s	17	53.70	59.44	51.85	42.59	Poor
DLOG_t12_3s	7	64.81	64.81	57.41	57.41	Poor
DLOG_t18_2s	17	53.70	62.22	50.00	35.19	Poor
DLOG_t18_3s	7	62.96	64.81	57.41	44.44	Poor
omnibus_DLOG_2s_2s	16	57.41	56.48	55.56	57.41	Poor
omnibus_DLOG_2s_3s	9	64.81	62.96	55.56	59.26	Poor
omnibus_DLOG_3s_2s	20	59.26	68.52	48.15	57.41	Poor
omnibus_DLOG_3s_3s	14	59.26	64.81	48.15	59.26	Poor
omnibus_2s_2s	20	88.89	90.74	87.04	81.48	Very Good
omnibus_2s_3s	10	90.74	92.59	90.74	87.04	Excellent
omnibus_3s_2s	18	88.89	87.04	85.19	83.33	Very Good
omnibus_3s_3s	11	90.74	90.74	88.89	87.04	Excellent
all_in_2s	20	90.74	87.04	85.19	83.33	Very Good
all_in_3s	11	90.74	92.59	90.74	87.04	Excellent

Table 3.10: Group 5 female candidates getting to assessment within 18 months of training, training model performance

Dataset	n	NB	SMO	IBk	J48	Rating
Psychosocial_2s	21	80.78	85.78	69.84	69.22	Good
Psychosocial_3s	11	85.16	83.44	67.03	70.31	Good
Training_2s	16	67.50	61.25	56.25	54.69	Poor
Training_3s	8	67.97	67.03	42.66	55.00	Modest
Consolidation_2s	24	83.28	74.06	85.62	73.28	Good
Consolidation_3s	11	78.44	82.03	81.72	78.91	Very Good
omnibus_survey_2s_2s	20	89.37	81.41	71.72	65.94	Good
omnibus_survey_2s_3s	11	86.88	80.78	76.72	76.25	Good
omnibus_survey_3s_2s	22	85.16	75.16	69.84	72.97	Good
omnibus_survey_3s_3s	13	84.06	81.25	86.56	62.50	Very Good
DLOG_t_2s	19	68.91	75.47	60.16	74.38	Good
DLOG_t_3s	12	68.28	74.37	62.19	70.47	Modest
DLOG_t6_2s	19	69.53	70.31	55.00	74.06	Modest
DLOG_t6_3s	13	69.06	72.50	60.00	69.53	Modest
DLOG_t12_2s	19	69.53	70.31	55.00	74.06	Modest
DLOG_t12_3s	13	69.06	72.50	60.00	69.53	Modest
DLOG_t18_2s	21	73.91	83.12	57.81	73.75	Good
DLOG_t18_3s	12	64.69	72.19	64.06	70.31	Modest
omnibus_DLOG_2s_2s	22	69.53	66.56	60.78	73.59	Modest
omnibus_DLOG_2s_3s	5	70.47	65.31	63.13	56.56	Modest
omnibus_DLOG_3s_2s	23	69.53	73.28	62.97	71.56	Good
omnibus_DLOG_3s_3s	16	66.72	68.44	68.44	71.72	Modest
omnibus_2s_2s	16	89.53	88.28	70.63	72.97	Very Good
omnibus_2s_3s	10	87.34	84.06	78.44	74.38	Very Good
omnibus_3s_2s	18	89.53	87.03	79.06	75.31	Very Good
omnibus_3s_3s	12	90.16	88.91	83.44	74.53	Very Good
all_in_2s	17	91.09	88.91	70.31	76.09	Very Good
all_in_3s	7	86.56	85.63	86.25	80.16	Very Good

Table 3.11: Group 5 passing first time, training model performance. Data standardised within sex

Dataset	n	NB	SMO	IBk	J48	Rating
Psychosocial_2s	18	63.04	56.74	65.22	67.39	Modest
Psychosocial_3s	9	60.87	58.04	54.35	36.96	Poor
Training_2s	18	52.17	54.35	58.70	41.30	Poor
Training_3s	9	56.52	56.96	47.83	30.43	Poor
Consolidation_2s	19	60.87	65.43	60.87	41.30	Modest
Consolidation_3s	10	63.04	65.22	63.04	56.52	Modest
omnibus_survey_2s_2s	19	65.22	60.87	71.74	47.83	Modest
omnibus_survey_2s_3s	8	73.91	67.83	82.61	60.87	Good
omnibus_survey_3s_2s	23	52.17	66.74	58.70	54.35	Poor
omnibus_survey_3s_3s	10	50.00	73.26	60.87	39.13	Poor
DLOG_t_2s	16	63.04	56.09	56.52	41.30	Poor
DLOG_t_3s	9	56.52	61.96	52.17	50.00	Poor
DLOG_a_2s	18	65.22	60.87	56.52	63.04	Modest
DLOG_a_3s	5	58.70	29.57	69.57	63.04	Modest
previous_courses_2s	8	58.70	63.48	58.70	58.70	Poor
previous_courses_3s	6	56.52	57.83	52.17	30.43	Poor
omnibus_DLOG_2s_2s	23	60.87	52.39	63.04	65.22	Modest
omnibus_DLOG_2s_2s	23	63.04	62.83	71.74	73.91	Modest
omnibus_DLOG_3s_2s	12	56.52	59.13	56.52	56.52	Poor
omnibus_DLOG_3s_3s	7	63.04	49.13	63.04	63.04	Modest
omnibus_2s_2s	22	65.22	69.35	71.74	47.83	Modest
omnibus_2s_2s	22	69.57	60.87	65.22	50.00	Modest
omnibus_3s_2s	24	60.87	48.04	58.70	54.35	Poor
omnibus_3s_3s	11	60.87	57.39	60.87	54.35	Poor
all_in_2s	18	76.09	69.57	82.61	60.87	Good
all_in_2s	18	54.35	45.65	67.39	60.87	Poor

3.7.2 Male getting to assessment training model performance

3.7.3 Female getting to assessment training model performance

3.7.4 First time pass model performance

Chapter 4

Self-Efficacy and Quality Mountain Days

We have finished a nice book.

Chapter 5

General Discussion

We have finished a nice book.

Appendix A

Developing the survey tool

A.1 Study 1 - Item selection

The aim of this study was to identify a suitable measure for each construct identified in Chapter 2 **LINK**, which could then be used to identify the most important variables for discriminating candidates who do complete the ML qualification from those who do not. Given that we wanted to collect data from candidates for 166 variables, using full-length measures of the relevant constructs would be unreasonable for participants. To measure each construct of interest with a full length measure would create a survey so long that few candidates would complete it and those that did would likely not be representative of the population. Horvath and Röthlin (2018) discuss additional issues associated with using "basic research questionnaires" that are not best suited to use in an applied sport psychology setting, where instead, one may only want to use a questionnaire as a screening tool, or help provide "a complete picture about an athlete's situation." As such we were particularly interested in identifying short-form measures as using such measures was most likely to allow us to create a suitably short survey to collect data with.

The development of short-form measures to reduce the burden on participants has been of interest to researchers for over 100 years (Smith et al., 2000). However, the development of short-form measures has attracted some criticism (e.g., Levy, 1968; Smith et al., 2000; Wechsler, 1967). One of the main criticisms of short-form measures has been that "rigorous, valid, comprehensive assessment is crucial for the evaluation and treatment of many psychological problems" (Smith et al., 2000, p 102) and that the time saving afforded by a short-form measure does not warrant the loss of validity associated with measuring a construct with fewer items.

When creating, or identifying, a short-form measure one should not assume that the evidence for the validity and reliability of the original measure applies to the short-from, therefore it is important to provide evidence for the reliability and validity of the short-form (Smith et al., 2000). This evidence should include, but is not limited to, reliability of the short-from, shared variance between the full- and short-form measure, content validity/coverage of the construct, and also that the reduction in items offers a meaningful reduction in the time taken for the measure to be completed (Horvath and Röthlin, 2018; Smith et al., 2000).

This project's aim was to identify the most important discriminatory variables for identifying candidates who do or do not complete the ML qualification using Machine Learning techniques rather than testing the relationships between variables using regression based techniques or structural equation modelling. Therefore, instead of using full-length measures to collect data for each construct, we used one or two item *indicators* for each construct, unless a suitable short-form measure existed (e.g., the Ten Item Personality Inventory Gosling et al., 2003).

A.1.1 Method

A.1.1.1 Procedure

Although using full measures was not a realistic aim in the project, we still felt that the reliability and validity of the indicators that we would use was paramount. Researchers have suggested a variety of ways in which short-form measures can be developed whilst remaining both reliable and valid. Considering the guidance provided by Smith et al. (2000) and Horvath and Röthlin (2018) along with the aim of this research, we used the steps below to identify items which would be used to collect data from candidates.

Our first choice was to identify existing suitable short-form measure. When this was not possible, but there was an existing measure that we were able to access secondary data for, we used the following steps. Firstly, we checked that existing measure did measure the construct of interest and that there was sufficient evidence for its reliability and validity. Secondly, we identified which items we wanted to retain based on both content validity and factor loadings. It was important that the items retained still provided adequate coverage of the construct. In some instances, this meant retaining an item which had a (relatively) low factor loading, but measured a unique aspect of that construct, was more important than it was to retain items with high factor loadings. This will necessarily have lowered the reliability coefficient for the short-form measure, however, it is important to note that internal consistency is only one aspect of valididty. If we did not feel that this process left us with items that provided adequate coverage of the construct of interest, we made a note of how the new, narrower, construct differed from the original.

Once we had identifed the items we wished to retain, we fitted a single factor latent variable model for both the full- and short-form measure to the secondary data, using `lavaan` (?), to estimate factor scores for each participant. These factor scores were then used to calculate a Pearson's correlation coefficient between predicted factor sores for full- and short-form measure as an estimate of shared variance. We believe that this method is better than correlating the item sum-scores as latent variables account for measurement error, thus, reducing the likelihood of receiving an optimistically biased estimate due to error correlation. Shared variance with the full measure was our main concern for this study as if the correlations are high enough then the two measures can be thought of as approximately equal (Smith et al., 2000). Finally, we calculated the composite reliability for the new short-form measure (ω; Fornell and Larcker, 1981)[1].

If secondary data were not available but we identified a suitable measure, we chose the best item(s) based on face validity of the items and factor loadings reported in the original paper validating the full measure. Finally, if none of the options above were possible, we developed item(s) within the research team in collaboration with Mountain Training UKI.

A.1.1.2 Measures

> (How much detail do I need to go into in the section - particularly for scales that already exist? Is it enough to say that there is a list of variables and measures in the supplementary information?) (Talk about domains, full information available in supplementary information.)

A.1.1.2.1 Personality measures

- To measure the "Big-Five" personality traits (openness, conscientiousness, extraversion, agreeableness and emotional stability), we used the Ten Item Personality Inventory (TIPI; Gosling et al., 2003). The TIPI comprises ten pairs of items (e.g., "Critical, quarrelsome"), one positively worded and one negatively worded for each trait. Each item has the same stem, "I see myself as…" Participants are then asked to score each item on a seven-point Likert scale from "Disagree strongly" (1) to "Agree strongly" (7) and sum scores are calculated for each of the five traits.
- Resilience and Robustness - Brief Resilience Scale (BRS; Smith et al., 2008)
- Perfectionism (Personal Standards and Concern Over Mistakes) - Frost Multidimensional Perfectionism Scale (FMPS; Frost et al., 1990)
  - **Do I want to include something about reanlaysing the YIPS study and finding similar results? What is the difference in classification rates/discriminant function?**
- Robustness of confidence - Trait Robustness of Self-Confidence Inventory (TROSCI; Beattie et al., 2011)

A.1.1.2.2 Socio-demographic

Some socio-demographic data are available on the CMS; however, some is not (e.g., income level, education level). To measure these, we used standard socio-demographic questions (e.g., "What is the highest level of school you had completed or the highest degree you had received when you registered?")

[1] Horvath and Röthlin (2018) highlight the fact that a measure of internal consistency such as Cronbach's alpha may be lower for a short-form measure simply because it has fewer items, therefore this may be a less useful metric than it would be in a longer measure. Composite reliability does not include the number of items in its calculation therefore is more suitable for assessing the internal consistency of short-form measures.

A.1.1.2.3 Self-efficacy scale

Perceived self-efficacy is domain specific and individuals will have varying levels of self-effiacy beliefs across different domains of their lives, therefore it is important that any measure of perceived self-efficacy is domain specific (Bandura, 1997, 2006). Mountain Training provide clear documentation about what will be required of candidates during their assessment, this includes a candidate handbook and syllabus (Mountain Training UK, 2015a), and a separate skills checklist (Mountain Training UK, 2015b). The first author conducted an inductive content analysis (Cho and Lee, 2014) of these documents to identify a list of skills, which should be able to perform on an ML assessment. This list of skills was then discussed with Mountain Training UKI who agreed that it provided good coverage of the skills that would be covered on an assessment.

Using the list of skills, a self-efficacy scale was created following Bandura's (2006) guidelines. The resultant scale was then piloted with Mountain Training UKI staff who provided feedback on the items, which was then collated and used to refine the scale. The final scale was made up of eleven items (e.g., "lead a group effectively in the mountains") rated on a scale of 0 (could not do at all) to 100 (highly certain could do) with a mid-point anchor (50; moderately could do). The items could then be presented to participants three times, each with a different introduction as we wanted to measure efficacy at two points along the pathway and also candidates' ideal efficacy levels:

1) Please rate how confident you were that you could do them immediately after your training course.
2) Please rate your degree of confidence, as of now/at your (first) assessment[2].
3) Now we know about your levels of confidence to perform these tasks as of now/at your (first) assessment, we would like to understand how confident you feel that your ideal self would be/have been at your (first) assessment. The Ideal Self: "Your ideal self is the kind of person you'd really like to be. It is defined by the characteristics you would ideally like to have. It's not necessary that you have these characteristics now, only that you believe you want to have them."

A.1.1.2.4 Personal projects

We used a modified version of Little's (1983) Personal Project Analysis, similar to that used by Beattie et al. (2015). The instructions were adapted and read:

> We are interested in studying the kinds of personal projects that candidates have at different stages of their life and how they relate to candidates' motivation to become an ML. All of us have a number of personal projects at any given time that we think about, plan for, and sometimes (though not always) complete.
>
> Please take a moment to think about the projects or goals that you were working on before your assessment, these may include things that you have already told us about.

[2]Different wording was presented to candidates based on whether or not they had been assessed.

Participants were then given examples of goals (e.g., "Completing another outdoor qualification," "Spending more time with my family") and asked to "write down the two goals that you were most likely to work towards in the six-months before your assessment, *not-including* becoming an ML." On the following page, for each of their stated goals and for the goal of "becoming an ML," they were then asked to rate the: importance of the goal, "not at all important to me" (0) to "extremely important to me" (100); progress towards the goal in the last six months/six months before their assessment, "no progress" (0) to "most progress" (100); and their perceived self-efficacy of attaining the goal, "I definitely do/did not have" (0) to "I definitely have/had" (100).

Using the scores provided, the following can then be calculated: relative importance, relative progress, and relative efficacy score using the following formula[3]:

$$Relative = \frac{Mountain\,Leader + 1}{(Goal\,1 + Goal\,2) \div 2 + 1}$$

A.1.1.2.5 Motives

In Chapter 2 it appeared that two different levels of motive were important to the completion of the ML qualification: participatory (the goal content) and regulatory (the "why"). To measure the participatory motives, we employed a similar methodology to Sheldon and Elliot's (1999) adaptation of Little's (1983) Personal Project Analysis. First, we asked participants to list two goals that they hoped to achieve by registering for the ML qualification. These reasons were then coded by the first author on a scale of definitely intrinsic (1) to definetly extrinsic (5) and a mean score was calculated. Examples for each value are as follows: (1) "To have fun," (2) "Being better equipped to enjoy the mountains safely for myself,"(3) "Assessing my own ability,"(4) "Confidence in leading groups in the moutnains,"(5) "Gain the ML qualification."

On the next page of the survey participants were asked to rate each reason they had given in terms of their behavioural regulation. Each item had the same stem, "I pursue this goal because…" The intrinsic item was "of the fun and enjoyment it provides me," the integrated reason was "it is a part of who I am or aspire to be," the identified reason was "I really believe it's an important goal to have," the interjected reason was "I would feel ashamed, guilty, or anxious if I didn't," the external reason was "someone else wants me to or because the situation demands it." Participants scored each of these reasons on a visual analogue scale with five equally spaced anchors from strongly agree (0) to strongly disagree (100), a mean score for each of the regulatory motives was then calculated.

A.1.1.2.6 Course staff coaching behaviours

The Military Coaching Behaviour Scale (MCBS; Wagstaff et al., 2018) is a 22-item scale that assesses five coaching behaviours: observing and performance analysis, effective questioning, goal setting, developmental feedback, and motivational feedback.

[3] To avoid returning an undefined value one is added to both the numerator and denominator.

A.1.1.2.7 Need supportive environment

The Perceived Environmental Supportiveness Scale (PESS; Markland and Tobin, 2010) measures autonomy support, structure, and involvement, each with five items.

A.1.1.2.8 Perceived conflict on courses

- Stuff from Matt
- Intra group conflict scale for sport (manuscript in prep Boulder, Hardy, Roberts, Woodman)

A.1.1.2.9 Social Support

- The Perceived Available Support in Sport Questionnaire (PASSQ; Freeman et al., 2011)
- The Athletes' Received Support Questionnaire (ARSQ; Freeman et al., 2014)

A.1.1.2.10 Preparation for assessment

Preparation for an assessment may encompass a variety of different things for different candidates and we were interested in how much candidates felt that they had done to prepare for an assessment. We asked participants to complete the sentence, "I have done ______ to prepare effectively for an ML assessment" using a visual analogue scale, anchored at "nothing" and "all that I could." Given the complex nature of this question, we first asked participants to list some of the things that they had done in the last six-months/six-months prior to their assessment to prepare. This was done to help improve accuracy in responses in a similar fashion to the decomposition questions used in The World Health Organization Health and Work Performance Questionnaire (Kessler et al., 2003; Means and Loftus, 1991).

A.1.1.2.11 Life events

Based on the results of Chapter 2, we wanted to measure change in three domains of candidates lives: social, professional, and health. The Recent Life Change Questionnaire (RLCQ; Miller and Rahe, 1997) has items covering these domains. At this point, we were not concerned about the exact events that may, or may not, have occurred, therefore we presented items from the RLCQ as examples for each domain and then asked participants to rate the extent to which they had experienced change in that domain of their life since their training course using a visual analogue scale (no change to major change). Another consideration when choosing this method was the sensitive nature of some life events. Allowing participants to indicate a magnitude of perceived change rather than explicitly responding to a sensitive item (e.g., "Miscarriage or abortion," "Being held in jail") was deemed more appropriate for this study.

Table A.1: Latent variable correlations between full- and short-form measures.

Measure	Reference	Variable	n	r	95%_CI	ω full	ω shor
FMPS	@Frost1990	Personal standards	120	0.81	0.74,0.87	0.80	0.5
		Concern over mistakes		0.75	0.66,0.82	0.74	0.5
BRS	@Smith2008	Resilience	192	0.97	0.96,0.98	0.91	0.8
TROSCI	@Beattie2011	Robustness of confidence	267	0.89	0.87,0.91	0.81	0.6
MCBS	@Wagstaff2018	Observation		0.96	0.95,0.97	0.96	0.8
		Effective questioning	263	0.93	0.92,0.95	0.96	0.7
		Goal setting		0.95	0.93,0.96	0.96	0.8
		Developmental feedback		0.95	0.94,0.96	0.98	0.8
		Motivational feedback		0.97	0.97,0.98	0.98	0.8
conflict	@Boulter	Relationship	384	0.80	0.76,0.83	0.88	N.
		Process		0.79	0.74,0.82	0.85	N.
PASSQ	@Freeman2011	Emotional		0.97	0.96,0.97	0.90	0.8
		Esteem		0.95	0.93,0.96	0.85	0.7
		Informational	219	0.90	0.87,0.92	0.82	0.7
		Tangible		0.83	0.78,0.86	0.82	0.6
ARSQ	@Freeman2014	Emotional		0.95	0.94,0.96	0.91	0.8
		Esteem		0.95	0.94,0.96	0.91	0.8
		Informational		0.94	0.92,0.95	0.91	0.7
		Tangible		0.95	0.93,0.96	0.92	0.8

A.1.1.2.12 Aspirations, intentions, and expectations

A.1.1.3 Data

A.1.2 Results

(Need to add PESS data if available - UK Sport/Dave Markland?)

- Validity
  - Variables with validity evidence from other studies 11
  - Variables with validity evidence from this study 61
  - Variables that were self-efficacy items constructed specifically for this project 59
  - Variables available on CMS 9
  - Variables that are total scores 11
  - **WHICH variables don't have any validity evidence??**

- FMPS correlations are high, but the internal consistency measure is low
  - Was this because we wanted to keep items in that captured part of the construct that other bits didn't?

Using the process described above, 197 items were removed, assuming 7 seconds per item (Qualtrics, 2019), this equates to a survey that is approximately 23 minutes shorter.

A.1.3 Discussion

Study 1 identified suitable short measures for each construct of interest. Whilst this study significantly reduced the number of items that participants would be required to answer, there were still too many to ask a single participant to answer. Including all of the items identified thus far would require candidates to spend approximately 21 minutes answering questions; participants would also be required to read the information sheet, transition between pages, etc. Therefore, Study 2 sought to reduce the number of constructs, by identifying those which were not important discriminatory variables.

A.2 Study 2 - Item reduction

Having selected appropriate items for the variables we wanted to collect data for, we were left with an item pool of 186, which was deemed too many to ask any individual participant as most would not complete the survey and those who did, would probably not be representative of the population. Therefore, we created four surveys, each of which contained a subset of the variables that we wanted to collect data for. Each variable was included in at least two of the surveys and each pairwise combination of variables was included in at least one survey. This was done to both collect as much data as possible and also to ensure that two-way interactions could be explored.

Figure A.1 shows a visual representation of the distribution of constructs between the four groups, a full list of variables and the groups that they were asked in can be found in **REF**.

> (Need to edit image to make it more simple.) (Have I explained the time point thing higher up?)

A.2.1 Method

A.2.1.1 Participants

In November 2018, we contacted all candidates trained between 2008 and 2016 ($n = 3794$). Each candidate had been randomly assigned to one of four groups

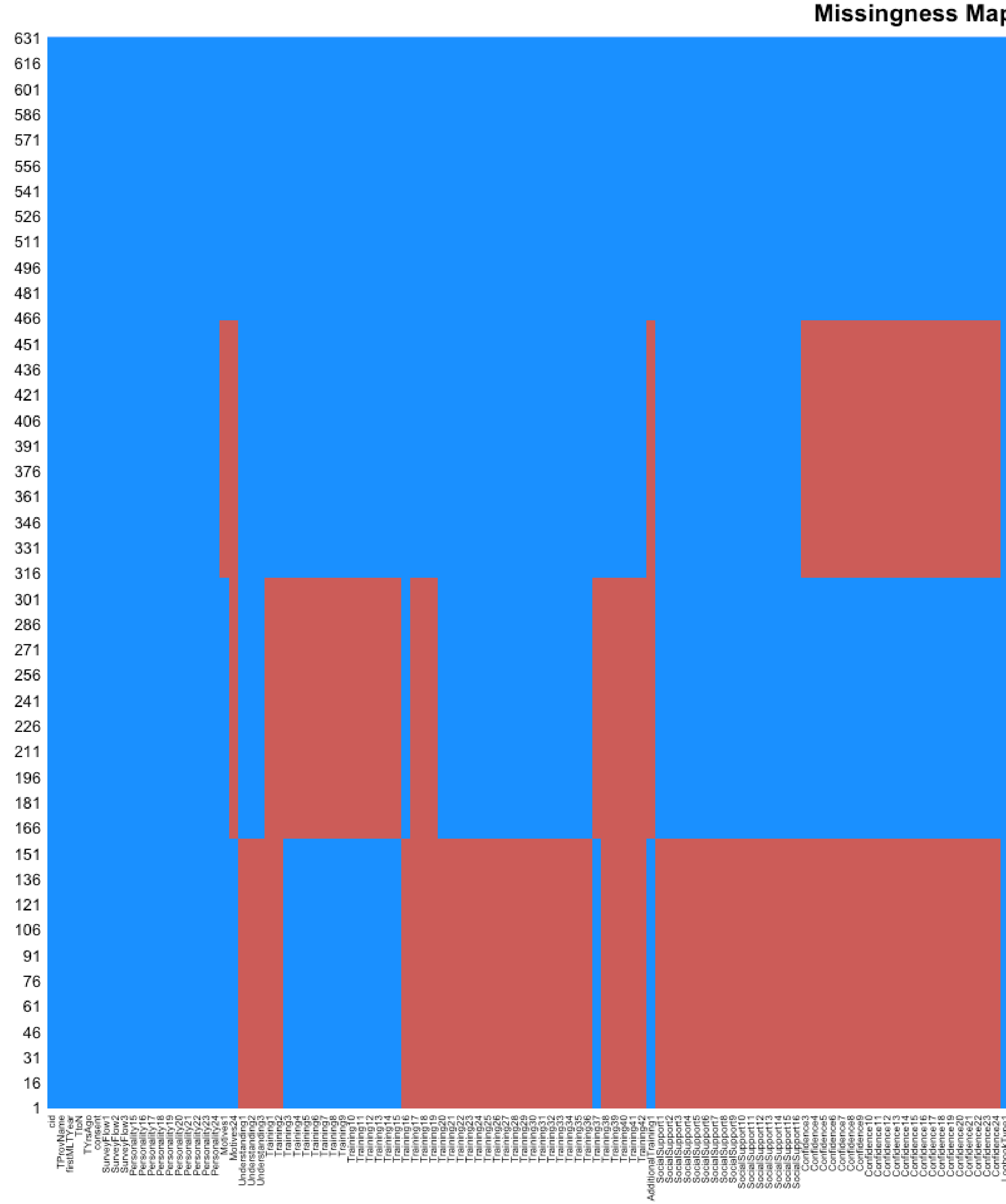

Figure A.1: Survey tool overlap

Table A.2: Survey participants per group

Group	n	% female	age	time
1	260	23.46	38.31	5.65
2	264	27.65	37.24	5.72
3	266	19.92	39.93	5.61
4	266	25.94	38.25	5.71

(stratified by year of training) using the `randomizr` package (**?**) and candidates from each of these groups were invited to complete one of the surveys described above using the Qualtrics survey platform (Qualtrics, 2019). Complete responses were collected from 1056 participants (Table A.2).

Once data collection was complete, each of the four groups was then split in two, one group for those candidates who had DLOG data and one group who did not have DLOG data. This was done as the pattern recognition procedure cannot handle missing data and we would then have had to omit all DLOG data, which would have left us unable to identify interactions between the survey and experience data. Once the groups had been split into these two groups, we created two data sets within each one for those who did not and then for each classification problem. This resulted in the following data sets for each survey group:

1) Getting to assessment within 18 months of training - no DLOG data
2) Getting to assessment within 18 months of training - with DLOG data
3) Passing the first assessment - no DLOG data
4) Passing the first assessment - with DLOG data

Figure A.2 provides a visual representation of the groups described above.

(need to manually position nodes)

In our data (and the population), most candidates have not been assessed 18 months after their training course. To ensure an orthogonal design (i.e., both groups were of equal size) we selected a random sample of candidates who had not been assessed 18 months after their training course of equal size to the group of candidates who had been assessed.

(Could/should I do something to check the representativeness of the samples (e.g., using sex, age, board)? Equivilance testing?)

A.2.1.2 Analytical method - pattern recognition

We used pattern recognition analyses to identify the most important discriminatory variables within each group; by identifying the most important, we were able to infer which variables were not important discriminatory variables. Pattern recognition analyses, originally developed in bioinformatics (Duda et al.,

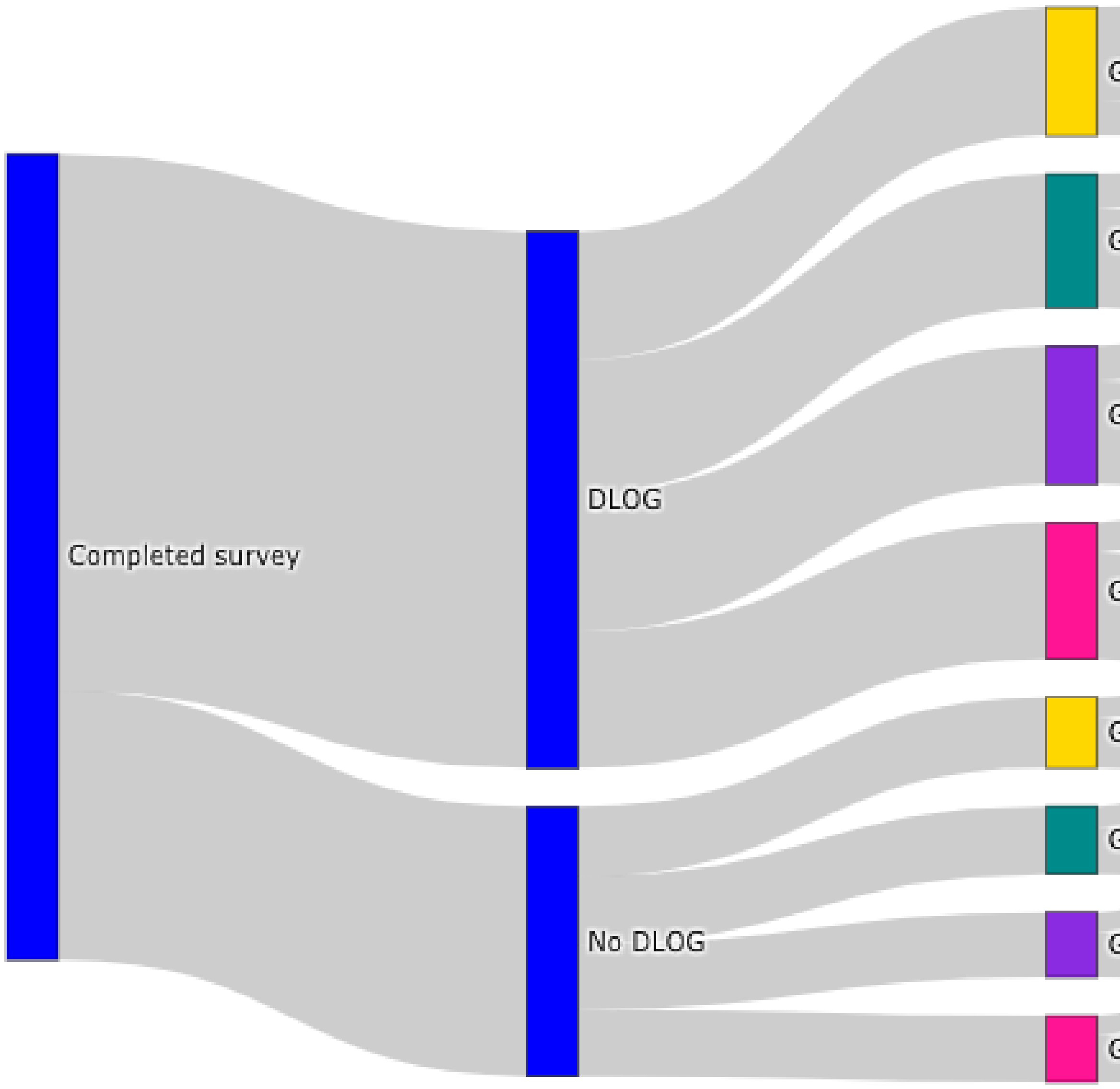

Figure A.2: Study 2 participants.

2000), use machine learning algorithms to identify a set of discriminatory features (variables), which can be used to identify the class of an object. Pattern recognition analysis is more appropriate for these data than "traditional" methods (e.g., discriminant function analyses) as they employ both linear and non-linear functions and therefore reflect multiple and complex interactions and not just "main-effects". Pattern recognition analyses have also been used in a number of recent studies to examine differences between athletes of different performance levels (e.g., Güllich et al., 2019; Jones et al., 2019).

More specifically, we used a pattern recognition procedure that has been developed for analysing what are known as *short and wide* data sets (i.e., data sets that contain more variables than cases; Jones et al., 2017). This procedure is a three-part process. First, we aimed to identify a set of features which correlated well with the class but had a low correlation with one another (feature selection). Then we tested the ability of this feature subset to correctly classify the candidates according to the criterion variable for that analysis (classification). Finally, we refined the feature subset to identify the simplest solution that best explained the data (recursive feature elimination).

We carried out the pattern recognition procedure outlined above twice for each of the four pilot surveys. The first set of analyses identified the most important features for discriminating candidates who get to assessment within 18 months of their training from those who do not. The second set of analyses identified the features which best discriminated candidates who passed their first assessment from those who did not.

Analyses were carried out using WEKA 3-9-3 open source software issued under the GNU General Public License version 3 (Bouckaert et al., 2018; Frank et al., 2016). WEKA is a machine learning workbench with a collection of algorithms widely used for data mining, machine learning, and pattern recognition.

> (Something about using ensembles (Bolón-Canedo et al., 2012, 2014) and also combination of supervised and unsupervised Is this still distributed feature selection but horizontally? Difference between individual and subset evaluation methods Tables 1 & 2 in Bolón-Canedo et al (2015b) have info about 3 of the 4 FS methods we use and their strengths & weaknesses)

A.2.1.2.1 Pre-processing

For feature selection and model identification, it is important that good quality data is used and that there are no missing values **REF**. Where there was missing data, we had a choice, remove the case or remove the variable[4]. The choice of removing the case or variable was decided by the amount of missing data within the variable in an effort to use as much data as possible. Where there were few cases with missing data for that variable, we removed the cases and where there were many cases with missing data for that variable, we removed the variable.

[4]Imputing values would not be appropriate for these data as this was usually due to candidate: (a) not having completed the survey, (b) having had an atypical journey through the pathway (e.g., received an exemption from training), or (c) provided an inappropriate answer for a question.

Once each of the 16 data sets had been cleaned, the data were standardised within sex using the `mousetrap` package (**?**), to control for sex differences[5].

> (How many variables and cases were excluded and what was the rationale?)

A.2.1.2.2 Feature Selection

Best practice guidelines for short and wide data recommend that feature selection is carried out using a number of different methods (Jones et al., 2017). With this in mind we used four feature selection algorithms: Correlation Feature Subset with a Best First Evaluator (CFS; Hall, 1999), Correlation Attribute Evaluator (CAE; Bouckaert et al., 2018), Relief-f (Kira and Rendell, 1992), and Support Vector Machine - Recursive Feature Elimination (SVM-RFE; Guyon et al., 2002). CAE, Relief-f, and SVM-RFE rank all variables in order of merit (magnitude of relationship), whereas CFS selects a subset of features that are highly correlated with the class but not with one another. As only CFS provides a subset of features, the top 20 features were selected from the rankings provided by the other three algorithms (as long as the attribute merit was greater than zero).

These analyses were carried out in a centralised fashion, with all variables being included at the same time (Bolón-Canedo et al., 2015a). We employed a *leave-one-out cross-validation* (LOO-CV) protocol, a special case of K-fold cross-validation, where $K = N$, as it reduces the impact of each object on the feature selection process by increasing the generalisability of the model (Hastie et al., 2009). Each data set was split into K parts or *folds*, with each part having an approximately equal number of cases. The Kth fold is then removed from the data and the feature selection algorithm is then applied to the remaining data, with each feature being assigned a merit score (or being selected/not for CFS), once this has been repeated K times the merit score for each attribute is averaged across the K iterations.

All of these are well established feature selection methods and the most important point to note about these four methods is that each works in very a different way. Therefore, the greater the number of algorithms which select a feature, the more confident one can be that it is important as it is less likely that the feature has been chosen by chance (Visa et al., 2011). For each of the 16 data sets this process yielded a number feature subsets, one containing those features which had been selected by at least two feature selection algorithms (2s), another containing those features which had been selected by at least three feature selection algorithms (3s), and where possible a third subset where features has been selected by all four algorithms (4s).

A.2.1.2.3 Initial Classification

For each of the feature subsets, we carried out *initial classification* experiments. Each of these experiments used four different classification algorithms with their

[5] We considered exploring the data having standardised them by training provider, however, we felt that the variance in number of candidates trained by each provider was too high.

default settings and LOO-CV: Naïve Bayes (NB; John and Langley, 1995), Sequential Minimal Optimization (SMO; Platt, 1998), Instance Based Learning (IBk; Aha et al., 1991), J48 Decision Tree (J48; Quinlan, 1993). As with feature selection, the more consistent the results from each algorithm (classification accuracy) for a feature subset, the more confidence we can place in the predictive validity of that subset.

An important point to note that as all of the data have been seen during the feature selection stage the classification rates may be slightly higher than they would be for previously unseen data (Kuncheva and Rodríguez, 2018; Smialowski et al., 2010). However, in this instance, the aim is not to create a predictive model, but to identify which features should be retained to create a survey tool and classification rates are only being used to compare the predictive power of feature subsets from the same data set - any inflation of classification rates will be shared by the feature subsets under consideration.

A.2.1.2.4 Final Classification

Having conducted the initial classification experiments, we sought to identify more parsimonious models, potentially with higher classification accuracies using the Recursive Elimination Method (Guyon et al., 2002). To do so we took each feature subset with more than five features in, examined the normalised SMO weight provided by the SMO classifier and removed the feature with the lowest weight before re-running the four classifiers on the, now smaller, feature subset. This continued in an iterative fashion until all features with a SMO weight $> .3$ had been removed, the iteration with the best classification rate was then retained.

A.2.2 Results

A.2.2.1 Classification rates

In 15 of the 16 datasets we were able identify a feature subset which could be used to correctly classify candidates with at least *good* accuracy (i.e., at least one classifier for that data set had a classification rate over 70%). For the other data set we were able to identify a feature subset which could be used to classify candidates with *moderate* accuracy (i.e., at least one classifier for that data set had a classification rate over 60%, Table 3.8).

Having identified a number of feature subsets that could be used to classify candidates in each group, we identified the items which we wanted to retain for the final survey. As not all items were asked to the same number of groups, we scored each item by the number of times that it was selected divided by the number of times it was asked. This was done so that the item retention process was not biased by the number of times that an item was asked. Items were retained if they were selected for the best models in at least half of the datasets they were asked to.

Of the 150 variables, 66 variables were retained based on the criteria above. Some of these variables were sum totals of constructs including variables not selected, therefore we chose to retain a further 23 variables. This process resulted in 134 items being retained for the final survey (see **APPENDIX** for a list of the variables retained for the final survey).

A.2.3 Discussion

This study sought to create a survey tool which could be administered to candidates who had completed an ML training course in order to help us identify the most important discriminatory variables for candidates who: (a) did or did not get to an assessment within 18 months of their training course and (b) did or did not pass their first assessment. Whilst no single candidate provided data for all of the variables, we were able to discriminate candidates with a degree of accuracy substantially greater than chance. The findings shows that firstly the measures used in the survey work, and secondly, that we collected data about variables which explain some of the variance in the criterion variables. It is important to note that just because a construct has not been selected, it is not necissarialy unimportant. Variables not selected as discriminatory variables may infact be important commonalities between the groups. The next study collected data from candidates who attended an ML training course between 2016 and 2018 on all of the variables retained following this study.

Including DLOG data in the models did not appear to improve the classification rates in any substantive way. This suggests that the variance explained by those data is better explained by survey variables. A likely explanation for this is that candidates use the DLOGs in different ways. Some candidates will log every experience that they have, some will log only the best of their experiences, some will log only their relevant experience, and some will log only the experience they need to meet the pre-requites for the course (potentially from an extremely large pool of experience). The use of DLOG in these different ways creates messy data, with no easy way to distinguish a candidate who only has 40 QMDs and a candidate who has far more than that but only logs 40 as they do not feel it would benefit them to log more.

> (Ross: This implies to me that if there were two items to measure a construct (for example) you put each item and the sum total of the construct into the analysis. If this is what you did (which I am pretty sure it is), I think its important that earlier you make this point clear and explain why it is appropriate/good.)

Bibliography

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1):37–66.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Armitage, C. J. and Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4):471–499.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2):122–147.

Bandura, A. (1997). *Self-efficacy*. Freeman, New York.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In Pajares, F. and Urdan, T. C., editors, *Self-efficacy beliefs of adolescents*, chapter 14, pages 307–337. Information Age Publishing, Greenwich.

Beattie, S., Hardy, L., Savage, J., Woodman, T., and Callow, N. (2011). Development and validation of a trait measure of robustness of self-confidence. *Psychology of Sport and Exercise*, 12(2):184–191.

Beattie, S., Hardy, L., and Woodman, T. (2015). A longitudinal examination of the interactive effects of goal importance and self-efficacy upon multiple life goal progress. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 47(3):201–206.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2012). An ensemble of filters and classifiers for microarray data classification. *Pattern Recognition*, 45(1):531–539.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2014). Data classification using an ensemble of filters. *Neurocomputing*, 135:13–20.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2015a). Distributed feature selection: An application to microarray data classification. *Applied Soft Computing Journal*, 30:136–150.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2015b). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems*, 86:33–45.

Bouckaert, R. R., Frank, E., Hall, M. A., Kirkby, R., Reutemann, P., Seewald, A., and Scuse, D. (2018). *WEKA Manual for Version 3-9-3*. University of Waikato, Hamilton.

Cho, J. Y. and Lee, E.-H. (2014). Reducing confusion about Grounded Theory and qualitative content analysis: Similarities and differences. *The Qualitative Report*, 19(32):1–20.

Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification*. John Wiley & Sons, 2nd editio edition.

Fornell, C. and Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1):39–50.

Frank, E., Hall, M. A., and Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, fourth edi edition.

Freeman, P., Coffee, P., Moll, T., Rees, T., and Sammy, N. (2014). The ARSQ: The Athletes' Received Support Questionnaire. *Journal of Sport and Exercise Psychology*, 36(2):189–202.

Freeman, P., Coffee, P., and Rees, T. (2011). The PASS-Q: the perceived available support in sport questionnaire. *Journal of sport & exercise psychology*, 33(1):54–74.

Frost, R. O., Marten, P., Lahart, C., and Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, 14(5):449–468.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528.

Gould, D. (2005). Goal setting for peak performance. In Williams, J., editor, *Applied sport psychology: Personal growth to peak performance*. Mayfield, Palo Alto, CA, fifth edit edition.

Güllich, A., Hardy, L., Kuncheva, L. I., Laing, S., Evans, L., Rees, T., Abernethy, B., Côté, J., Warr, C., and Wraith, L. (2019). Developmental Biographies of Olympic Super- Elite and Elite Athletes: A Multidisciplinary Pattern Recognition Analysis. *Journal of Expertise*, 2(1):23–46.

Guyon, I., Weston, J., and Barnhill, S. (2002). Gene selection for cancer classification using Support Vector Machines. *Machine Learning2*, 46:389–422.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. Doctoral thesis, University of Waikato.

Hardy, L., Barlow, M., Evans, L., Rees, T., Woodman, T., and Warr, C. (2017). Great British medalists: Psychosocial biographies of Super-Elite and Elite athletes from Olympic sports. *Progress in Brain Research*, 232:1–119.

Hardy, L., Jones, G., and Gould, D. (1996). *Understanding psychological preparation for sport*. John Wiley & Sons, Chichester.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. *Bayesian Forecasting and Dynamic Models*, 1:1–694.

Horvath, S. and Röthlin, P. (2018). How to Improve Athletes' Return of Investment: Shortening Questionnaires in the Applied Sport Psychology Setting. *Journal of Applied Sport Psychology*, 30(2):241–248.

John, G. H. and Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo.

Jones, B. D. (2019). *'Game Changers': Discriminating Features within the Microstructure of Practice and Developmental Histories of Super-Elite Cricketers - A Pattern Recognition Approach*. Phd, Bangor University.

Jones, B. D., Hardy, L., and Kuncheva, L. I. (2017). Pattern Recognition Analysis: Procedures for SSHES - Stage 1 WEKA.

Jones, B. D., Hardy, L., Lawrence, G. P., Kuncheva, L. I., Du Preez, T. L., Brandon, R., Such, P., and Bobat, M. (2019). The identification of 'Game Changers' in England cricket's developmental pathway for elite spin bowling: a machine learning approach. *Journal of Expertise*, 2(2):92–120.

Judge, T. A., Bono, J. E., Ilies, R., and Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4):765–780.

Kessler, R. C., Barber, C., Beck, A., Berglund, P., Cleary, P. D., McKenas, D., Pronk, N., Simon, G., Stang, P., Ustun, T. B., and Wang, P. (2003). The World Health Organization Health and Work Performance Questionnaire (HPQ). *Journal of Occupational and Environmental Medicine*, 45(2):156–174.

Kira, K. and Rendell, L. A. (1992). Feature selection problem: traditional methods and a new algorithm. *Proceedings Tenth National Conference on Artificial Intelligence*, pages 129–134.

Kuncheva, L. I. and Rodríguez, J. J. (2018). On feature selection protocols for very low-sample-size data. *Pattern Recognition*, 81:660–673.

Levy, P. (1968). Short-form tests: A methodological review. *Psychological Bulletin*, 69(6):410–416.

Little, B. R. (1983). Personal projects: A rationale and method for investigation. *Environment and Behavior*, 15(3):273–309.

Markland, D. and Tobin, V. J. (2010). Need support and behavioural regulations for exercise among exercise referral scheme clients: The mediating role of psychological need satisfaction. *Psychology of Sport and Exercise*, 11(2):91–99.

Means, B. and Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, 5(4):297–318.

Miller, M. A. and Rahe, R. H. (1997). Life changes scaling for the 1990s. *Journal of Psychosomatic Research*, 43(3):279–292.

Mountain Training (2019). Mountain Leader FAQs.

Mountain Training UK (2015a). Mountain Leader - Candidate Handbook. pages 1–26.

Mountain Training UK (2015b). Mountain Leader Award: Skills checklist.

Platt, J. C. (1998). Fast training of Support Vector Machines using Sequential Minimal Optimization. In Schoelkopf, B., Burges, C., and Smola, A., editors, *Advances in Kernel Methods - Support Vector Learning*.

Qualtrics (2019). Qualtrics.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., Montgomery, H., Laing, S., and Warr, C. (2016). The Great British Medallists Project: A Review of Current Knowledge on the Development of the World's Best Sporting Talent. *Sports Medicine*, 46:1041–1058.

Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research*, 3:1371–1382.

Roberts, R., Woodman, T., Hardy, L., Davis, L., and Wallace, H. M. (2013). Psychological Skills Do Not Always Help Performance: The Moderating Role of Narcissism. *Journal of Applied Sport Psychology*, 25(3):316–325.

Seery, M. D. and Quinton, W. J. (2016). Understanding Resilience: From Negative Life Events to Everyday Stressors. In *Advances in Experimental Social Psychology*.

Sheldon, K. M. and Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model. *Journal of Personality and Social Psychology*, 76(3):482–497.

Smialowski, P., Frishman, D., and Kramer, S. (2010). Pitfalls of supervised feature selection. *Bioinformatics*, 26(3):440–443.

Smith, B. and McGannon, K. R. (2018). Developing rigor in qualitative research: problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*, 11(1):101–121.

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., and Bernard, J. (2008). The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine*, 15(3):194–200.

Smith, G. T., McCarthy, D. M., and Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1):102–111.

van de Schoot, R., Yerkes, M. A., Mouw, J. M., and Sonneveld, H. (2013). What Took Them So Long? Explaining PhD Delays among Doctoral Candidates. *PLoS ONE*, 8(7).

Visa, S., Ramsay, B., Ralescu, A., and van der Knapp, E. (2011). Confusion Matrix-based Feature Selection. In *23rd MAI & CS Conference*, Cincinnati.

Wagstaff, C. R. D., Arthur, C. A., and Hardy, L. (2018). The Development and Initial Validation of a Measure of Coaching Behaviors in a Sample of Army Recruits. *Journal of Applied Sport Psychology*, 30(3):341–357.

Waite, L. J. (1995). Does marriage matter? *Demography*, 32(4):483–507.

Webster, L. V., Hardy, J., and Hardy, L. (2017). Big Hitters: Important Factors Characterizing Team Effectiveness in Professional Cricket. *Frontiers in Psychology*, 8(July):1–18.

Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. Psychological Corporation, New York.

Weinberg, R. S. and Gould, D. (2014). *Foundations of sport and exercise psychology*. Human Kinetics, Champaign, IL, sixth edit edition.

Woodman, T., Zourbanos, N., Hardy, L., Beattie, S., and McQuillan, A. (2010). Do Performance Strategies Moderate the Relationship Between Personality and Training Behaviors? An Exploratory Study. *Journal of Applied Sport Psychology*, 22(2):183–197.

Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: a fast correlation-based filter solution. In *Proceedings of the twentieth international conference on machine learning*, pages 865–863.