

# Hur testar man Isokai-testet?

## Bedömningen av ett teoretiskt ramverk för att möjliggöra utvärdering av selekteringstest baserade på fysiska attribut

Adam Hofstedt Fröberg & André Stengård

### Abstract

The Swedish Armed Forces is one of Sweden's largest employers. Despite the organization having a stated goal to increase structural gender equality, women only constituted 8% of the organization's full-time military personnel at the end of 2016. One probable culprit is a test in the organization's recruitment process, the Isokai-test. All military applicants are required to pass this test, which aims to measure an applicant's overall strength. Of the female applicants that failed the recruitment process between 2006 and 2008, an average of 58% did so because of the Isokai-test. The corresponding value for men was 1%.

The use of a test that disqualifies one gender to such an extent needs to be properly justified with an academically substantiated approach. The Swedish Armed Forces have not done so. Only two suitable theoretical frameworks for evaluating physical employment tests (such as the Isokai-test) have been identified. The more comprehensive of these two are presented in the article *A framework for the design and development of physical employment tests and standards* (Payne & Harvey 2010). The intended purpose of this framework is to guide in the design of new physical employment tests. The objective of this study is to assess the usability of Payne & Harvey's framework for evaluating already existing physical employment tests. The study achieves this through qualitative interviews with individuals who have performed the Isokai-test and subsequently worked in the Swedish Armed Forces.

The analysis indicates that all theoretical concepts presented in Payne & Harvey's framework are useful for evaluating physical employment tests. However, the support for this conclusion is limited and the relative usability of the different concepts could not be established. In addition, two empirically important phenomena are observed which are not identified in Payne & Harvey's framework. The first phenomenon is the distinction between current and potential strength, the second is the monetary cost of a physical employment test. In any future evaluation of a physical employment test it is advised that both phenomena are taken into consideration.

---

**Keywords:** armed forces, physical employment test, selection, human resource management

**Handledare:** Laurence Romani  
**Kursansvarig:** Laurence Romani

**Framläggning:** maj 2017

# **Vi vill tacka alla er som har hjälpt och stöttat oss under denna vår**

Vår handledare

**Laurence**

Vår handledningsgrupp

**Eleonor**

**Elin**

**Eliza**

**Mohammed**

**Rebecca**

**Sofia**

Vänner som givit värdefull kritik

**Filip**

**Denise**

**Vi är enormt tacksamma för ert engagemang**

Adam & André

2017-05-15

# Innehållsförteckning

Begrepp och definitioner .....	5
<b>1. Bakgrund .....</b>	<b>6</b>
1.1 Introduktion .....	6
1.2 Problemformulering .....	6
1.3 Payne och Harveys teoretiska ramverk .....	7
1.4 Studiens syfte och forskningsfråga .....	8
1.5 Selektionsprocessen och Isokai-testet .....	8
1.6 Studiens koppling till <i>management</i> som akademisk disciplin .....	9
1.7 Avgränsningar .....	9
<b>2. Teoretisk referensram .....</b>	<b>10</b>
2.1 Översikt av teori inom selektion .....	10
2.2 Selektion baserad på fysiska attribut .....	11
2.3 A framework for the design and development of physical employment tests and standards	12
2.3.1 Artikelöversikt .....	12
2.3.2 Summering av viktiga faktorer vid skapandet av ett fysiskt selektionstest.....	13
2.3.3 Kategorisering av fysiska selektionstest.....	17
2.3.4 Relationen mellan prestation på testet och prestation på jobbet .....	18
2.4 Problematiskering av begreppet validitet .....	18
<b>3. Metod .....</b>	<b>19</b>
<b>3.1. Metodval och forskningsansats .....</b>	<b>19</b>
3.1.1 Val av ansats.....	19
3.1.2 Forskningsstrategi .....	20
3.1.3 Forskningsdesign .....	20
<b>3.2 Urval.....</b>	<b>21</b>
3.2.1 Val av organisation .....	21
3.2.2 Val av intervjupersoner .....	21
<b>3.3 Kvalitativ undersökning .....</b>	<b>22</b>
3.3.1 Semistrukturerade djupintervjuer.....	22
3.3.2 Genomförande av intervjuer .....	23
3.3.3 Anonymisering .....	23
3.3.4 Metod vid bearbetning av data .....	23
<b>3.4 Metoddiskussion .....</b>	<b>23</b>
<b>4. Empiri.....</b>	<b>24</b>
<b>4.1 Diskussionstrender .....</b>	<b>24</b>
4.1.1 Kopplingen mellan Isokai-testet och de efterföljande fysiska arbetsuppgifterna.....	24
4.1.2 Isokai-testets mätbarhet och reliabilitet .....	25
4.1.3 Diskriminering genom testdesign och kravnivå.....	26
4.1.4 Isokai-testets praktiska genomförbarhet .....	26
4.1.5 Diskussionen kring alternativa test .....	27
4.1.6 Testandet av nuvarande styrka eller potential styrka .....	28
4.1.7 Indirekta kostnader .....	28

4.1.8 Övrig empiri av intresse för analysen .....	29
<b>5. Analys och diskussion .....</b>	<b>30</b>
5.1 Användbarheten hos de olika delarna i Payne och Harveys ramverk .....	30
5.2 Fenomen uppmärksammade vid appliceringen av Payne och Harveys ramverk på Isokai- testet som inte är en del av ramverket .....	32
5.2.1 Nuvarande eller potential styrka .....	33
5.2.2 Ett selekteringstests kostnad .....	33
<b>6. Slutsatser .....</b>	<b>34</b>
6.1 Besvarande av forskningsfråga 1 .....	34
6.2 Besvarande av forskningsfråga 2 .....	35
6.3 Praktiska implikationer .....	36
6.4 Studiens begränsningar .....	36
6.5 Förslag till vidare forskning .....	36
<b>7. Referenser .....</b>	<b>38</b>
7.1 Digitala källor .....	38
7.2 Tryckta källor .....	40
<b>8. Bilagor .....</b>	<b>45</b>
8.1 Bilaga 1: Intervjumall .....	45
8.2 Bilaga 2: Förteckning över figurer och tabeller .....	47

## Begrepp och definitioner

Antagningsprövning	En samling tester, undersökningar och samtal över en eller två dagar för att testa den sökandes lämplighet för Försvarsmakten och dess olika befattningar (Rekryteringsmyndigheten 2016a, 2017a).
Grundläggande militär utbildning (GMU)	En initial, tre månader lång militär utbildning som ska vara likadan oavsett utbildningsort eller sökt befattning (Försvarsmakten 2017a).
Befattningsutbildning	Fortsatt utbildning inom det område (befattning) den sökande blivit antagen till. Genomförs av soldater, sjömän och gruppbefäl. Utbildningens längd varierar mellan 1-8 månader (Försvarsmakten 2017b).
Förberedande officerskurs (FOK)	Ersätter befattningsutbildningen för vissa specialistofficers- och officersutbildningar. Utbildningens längd är cirka 4 månader (Försvarsmakten 2017c).
Grundutbildning	GMU samt antingen befattningsutbildning eller förberedande officerskurs (Försvarsmakten 2017d).
Isokai-testet	Den delen av antagningsprövningen som ska pröva den sökandes generella muskelstyrka, kallas för "muskelstyrketestet" av Rekryteringsmyndigheten (Rekryteringsmyndigheten 2010). Genomförs i en Isokai-maskin (Engelbrektsson 2011; Försvarsmakten 2015).
Isokai	Isokai är ett isokinetiskt träningsredskap som består av en stång fäst i ställinor (Engelbrektsson 2011).
Fysiskt selekteringstest	Ett begrepp introducerat i denna studie. Används synonymt med termen "selekteringstest baserat på fysiska attribut". Refererar till ett test där en individs fysiska förmåga utvärderas. Synonymt med begreppet <i>physical employment test</i> som används av Payne och Harvey (2010) och som definierats av Jackson & Hodgdon (2000). Begreppet selekteringstest kan även ses som synonymt med ordet urvalstest samt ordet utslagstest.
Diskriminering	I studien används begreppet diskriminering enligt Nationalencyklopedins definition för hur det används i dagligt tal: "I dagligt tal avser man med diskriminering den negativa formen, dvs. sådan särbehandling av en grupp eller av en individ som innebär olägenhet av något slag" (Bergman 2017). Ej att förväxla med <i>discriminatory power</i> som definierat av Payne och Harvey (2010).

Tabell 1 - Viktiga begrepp och dess definitioner.

# 1. Bakgrund

## 1.1 Introduktion

Försvarmakten är den myndighet och organisation som ansvarar för landets militära försvar (Rönnerberg 2017a) och fyller därför en betydande funktion i det svenska samhället. Organisationen har drygt 50 000 anställda varav cirka 20 000 är kontinuerligt tjänstgörande (Försvarmakten 2017e) och är därför en av Sveriges största arbetsgivare.

Försvarmakten har som uttalat mål att öka den strukturella jämställdheten i organisationen (Försvarmakten 2017f). Trots det utgjorde kvinnor vid slutet av 2016 enbart 8 % av kontinuerligt tjänstgörande militär personal och 6 % av den tidvis tjänstgörande (Försvarmakten 2017f).

Varför den strukturella jämställdheten är låg kan ha många orsaker. En av dessa orsaker är antagligen ett deltest inom selekteringsprocessen som mäter muskelstyrka, Isokai-testet, ett test man måste få godkänt resultat på för att få verka inom Försvarmakten (Försvarmakten - Högkvarteret 2015). Testet har tidigare fått kritik för att det förhindrar kvinnor från att påbörja en militär karriär inom Försvarmakten. Susanne Nyberg, ledamot i Officersförbundets förbundsstyrelse uttryckte i Officerstidningen år 2012 att:

*“Skulle jag få önska så skulle kraven vara mer differentierade. Där jag arbetar, inom flottan, ser vi till exempel inte någon direkt koppling mellan extrem muskelstyrka och vår verksamhet.”* (Langlet 2012).

Faktum är att statistik från Rekryteringsmyndigheten visar att av de kvinnor som inte klarade av selekteringsprocessen mellan 2006 och 2008 blev i genomsnitt 58 % underkända på Isokai-testet som delmoment. Motsvarande siffra för män är 1 % (Rekryteringsmyndigheten 2007, 2008, 2009). Dessa siffror är äldre men Rekryteringsmyndigheten publicerar ej statistik för senare år på sin hemsida. En statlig offentlig utredning (Christensen et al. 2016) redovisar att under 2015 är motsvarande siffror 9 % för kvinnor och 0 % för män. Reduceringen för kvinnor är markant men vi har inte kunnat finna någon förklaring till orsakerna bakom reduktionen eller om siffrorna kan komma att stiga. Skillnaden i procent mellan män och kvinnor är dock fortsatt stor och den offentliga utredningen refererar till det som ett problem (Christensen et al. 2016).

## 1.2 Problemformulering

På grund av att Isokai-testet förhindrar kvinnors inträde i Försvarmakten i den utsträckning det gör är det av vikt för organisationen att undersöka om användningen av Isokai-testet som utslagstest är välmotiverat. För att undersöka om Isokai-testet är ett välmotiverat test

behövs ett systematiskt tillvägagångssätt som tar sin grund i forskning kring selekteringsteori.

Under förarbetet till denna studie har vi inte kunnat finna några indikationer på att en sådan undersökning har genomförts gällande Isokai-testet. Utöver det fastställs i en vetenskaplig rapport från Karolinska institutet gällande Isokai-testet att:

*“[...] det i Försvarsmakten råder brist på underlag av olika arbetskrav och objektiva kravnivåer för olika befattningar. Det medför att arbetet med att utveckla olika tester för urval begränsas och att prestationsnivåer mot olika befattningar inte kan fastställas. Analyser om vad ett visst uppmätt värde av Isokai-testet innebär för reell förmåga i relation till olika befattningskrav har inte genomförts”* (Larsson et al. 2011).

Vidare nämner den offentliga utredningen som vi tidigare hänvisade till i sin diskussion angående Isokai-testet att det finns ett behov av att *“[...]säkerhetsställa att man mäter rätt förmågor utifrån befattningskrav.”*(Christensen et al. 2016).

Ovanstående utdrag tyder på att Isokai-testet användning som utslagstest inte har motiverats väl och att detta utgör ett problem.

Vid genomgång av relevant teori som skulle kunna användas för att utvärdera Försvarsmaktens användning av Isokai-testet kunde dock endast två teoretiska ramverk som berör selekteringstest baserade på fysiska attribut identifieras. Det mer genomgående och omfattande av dessa två ramverk introducerades av Payne och Harvey (2010) i artikeln *A framework for the design and development of physical employment tests and standard.*

### **1.3 Payne och Harveys teoretiska ramverk**

Syftet med ramverket (Payne & Harvey 2010) är att vägleda i konstruktionen eller valet av ett eller flera *physical employment tests*, med andra ord selekteringstest baserade på fysiska attribut. Således är artikeln inte utformad med syftet att vägleda i utvärderingen av redan existerande fysiska selekteringstest.

Artikeln består av tre större delar. Del ett summerar de faktorer som bör uppmärksammas vid valet eller konstruktionen av ett eller flera fysiska selekteringstest. Del två klassificerar fysiska selekteringstest längs ett kontinuum och del tre fokuserar på relationen mellan prestationen på testet och senare prestation på arbetsuppgifter. Artikeln och dess ramverk introduceras i detalj under avsnitt 2.3.

## 1.4 Studiens syfte och forskningsfråga

Syftet med studien är att bedöma användbarheten i Payne och Harveys ramverk vid utvärdering av redan existerande selekteringstest baserade på fysiska attribut. Studien ämnar således bidra till den akademiska diskussionen kring fysiska selekteringstest och utvärderingen av dessa.

Användbarheten i ramverket bedöms genom att applicera det på Försvarmaktens existerande Isokai-test med hjälp av en kvalitativ fallstudie. Fallstudien består av kvalitativa intervjuer med militär personal som har genomfört Isokai-testet för att sedan verka inom Försvarmakten. Genom att bedöma användbarheten i ramverkets enskilda aspekter, vid utvärderingen av ett redan existerande selekteringstest, avser vi bidra med insikter angående befintlig selekteringsteori.

För att tydliggöra, syftet är således inte att utvärdera Isokai-testet lämplighet, utan att bedöma huruvida Payne och Harveys ramverk kan användas för att utvärdera fysiska selekteringstest såsom Isokai-testet.

Med hänsyn till syftet presenteras följande forskningsfrågor:

1. *Vilka delar i Payne och Harveys ramverk är användbara vid utvärderingen av Försvarmaktens fysiska selekteringstest, Isokai-testet?*
2. *Vilka fenomen uppmärksammade vid appliceringen av Payne och Harveys ramverk på Försvarmaktens fysiska selekteringstest, Isokai-testet, är inte en del av ramverket?*

## 1.5 Selekeringsprocessen och Isokai-testet

Idag samarbetar Försvarmakten med Rekryteringsmyndigheten (tidigare Pliktverket) för att selektera militär personal (Rekryteringsmyndigheten 2016b). Processen för att ansöka som frivillig till Försvarmakten är väl definierad (Rekryteringsmyndigheten 2016a). Ansökan görs på Rekryteringsmyndighetens hemsida. På hemsidan får man därefter fylla i ett "Rekrytttest" där man svarar på frågor angående ens lämplighet. Går man vidare efter rekrytttestet kallas man till en "Antagningsprövning" under en eller två dagar där man genomgår: teoretiska prov, förundersökning av sjuksköterska, medicinska tester, samtal med psykolog samt fysiska tester (Rekryteringsmyndigheten 2016a). Blir man inte godkänd på någon av dessa tester eller deltester får man ej gå vidare och genomföra GMU (Försvarmakten - Högkvarteret 2015). Man får med andra ord inte utbilda sig till eller ta anställning som militär i Försvarmakten. Efter godkänd antagningsprövning följer samtal med yrkesvägledare och därefter grundläggande militär utbildning (Rekryteringsmyndigheten 2017a).

De fysiska testerna består av ett fysiskt arbetsprov på cykel samt en registrering av längd vikt och muskelstyrka (Rekryteringsmyndigheten 2017a). Testet av muskelstyrka kallas av Försvarmakten för "Muskelstyrketestet" men kallas i denna uppsats för Isokai-testet. Om man inte blir godkänd på Isokai-testet, eller vill ha ett högre resultat, kan man genomföra testet igen efter en spärrtid på tre månader (Försvarmakten - Högkvarteret 2015).

Isokai-testet består av ett sökande genomför ett så kallat "Långt drag" i en Isokai-maskin (Rekryteringsmyndigheten 2010; Engelbrektsson 2011). Ett långt drag består av ett marklyft för att i stående position direkt övergå i ett drag till hakan (Engelbrektsson 2011). Isokai-maskinen består av en stång som är fäst i ställinor vilket möjliggör för stången att röra sig horisontellt under draget. Maskinen bromsar hastigheten i draget till en konstant hastighet genom hela utförandet. Ju snabbare en sökande försöker dra, desto hårdare bromsar maskinen rörelsen och desto mer kraft registreras på en våg (tryckgivare) som den sökande står på (Larsson et al. 2010). Den genomsnittliga kraften genom hela draget registreras i Newton av en dator kopplad till Isokai-maskinen. Detta värde översätts till ett resultat från ett till nio, där 463N eller lägre ger en etta och 869N eller högre resulterar i en nia (Larsson et al. 2010).

## **1.6 Studiens koppling till *management* som akademisk disciplin**

En del av *human resource management* är selekteringen av en organisations personal (Dessler 2013). Praktiserande av effektiv *human resource management* har visat sig resultera i bättre finansiella resultat för organisationer (Huselid 1995). Att selektera rätt personal har också specifikt visat sig vara av vikt eftersom selekteringen påverkar organisationens prestation, kostnader samt rättsliga skyldigheter (Dessler 2013). Enligt Dessler (2013) är det huvudsakliga målet med selektering att uppnå *person-job fit*, det vill säga att få den potentiella personalens kunskap, kompetens och förmågor att överensstämma med de motsvarande krav som är avgörande för att prestera i yrket.

## **1.7 Avgränsningar**

Studien fokuserar på att bedöma användbarheten av Payne och Harveys teoretiska ramverk för att utvärdera Försvarmaktens fysiska selekteringstest, Isokai-testet, men ämnar inte utvärdera utformningen och lämpligheten av testet i sig.

Rekryteringsprocessen till Försvarmakten består av en mängd tester men denna studie berör enbart Isokai-testet.

Gällande selekteringsteori avgränsas studien till en bedömning av Payne och Harveys ramverk. Således hanterar studien inga andra ramverk inom området då vi anser att detta

ramverk är det mest relevanta, omfattande och uttömmande ramverket i relation till Isokai-testet.

## 2. Teoretisk referensram

### 2.1 Översikt av teori inom selektering

Som etablerats i avsnitt 1.6 är selektering av personal en viktig del av *human resource management*. *Human resource management* och i synnerhet selektering är dock ett forskningsområde som involverar flera akademiska discipliner. En stor del av det akademiska arbete som utförts inom selektering hamnar inom ett område som vanligtvis benämns *industrial and organizational psychology* och en stor del av de forskare som gjort akademiska bidrag till frågan om selektering är psykologer (Schmitt 2012).

Idag råder en utbredd teoretisk och praktisk konsensus kring den grundläggande process som organisationer ska använda vid skapandet och valideringen av selekteringsmetoder (Boxall, Purcell & Wright 2008). Denna process består av tre generella steg. Det första steget är *job analysis*. *Job analysis* ämnar definiera de uppgifter ett visst jobb innebär och den kunskap, kompetens, förmågor och andra karakteristiska (summerat *KSAO*) som krävs för att utföra dessa uppgifter på bästa sätt (bristen på *job analysis* i Försvarsmakten är vad Larsson et al. [2011] uppmärksammar i avsnitt 1.2). I det andra steget i processen tar selekteringsexperter fram mått eller indikationer på dessa *KSAO* som de kommer använda för att utvärdera potentiell personal med. Därefter sker själva rekryteringen. Det tredje steget i processen handlar om utvärdering. I detta steg bedömer man validiteten i den selekteringsmetod man skapat genom att mäta korrelationen mellan de i testet uppmätta värdena för de som anställdes och deras senare arbetsprestation (Guion 1998; Schmitt & Chan 1998).

Att teoretisk konsensus går att finna kring de större dragen i en selekteringsprocess innebär inte att konsensus råder angående andra delar av selekteringsprocessen eller de tre olika delprocesserna. Ett av de första konstateranden som görs i denna uppsats är vikten av *human resource management* och selektering för en organisations prestation (se avsnitt 1.6). Även i denna fråga pågår teoretiska debatten. Även fast den tidigare nämnda Huselid (1995) får stöd av andra forskare som Terpstra och Rozell (1993) samt Delery och Doty (1996) saknas inte heller kritiker. Boxall och Purcell (2000) samt Wright et al. (2001) kritiserar metoderna som använts vid genomförande av dessa tidigare nämnda studier. Kritiken handlar bland annat om att olika organisationers praktisering av HR-metoder är svårt att observera, med senior personal som uttalar sig inom områden de inte har ordentlig kunskap om och intervjufrågor som har en för ytlig karaktär. Wright et al. (2005) konstaterar också att användandet av HR-metoder i organisationer är starkt korrelerat med, inte bara framtida prestation, utan även med historisk prestation hos organisationen. Detta indikerar

att användandet av HR-metoder inte nödvändigtvis orsakar (positiva) prestationer hos organisationer utan att kausaliteten kan vara den omvända, eller att båda faktorer påverkas av andra, externa variabler.

Det existerar även andra stora diskussionsområden inom nutida selekteringsteori. En av de större är frågan om vad som egentligen utgör arbetsprestation av en anställd inom ett yrke. Diskussionen fokuserar på om arbetsprestation endast utgör de tekniska och direkt uppgiftsrelaterade prestationerna (Austin & Villanova 1992), eller om andra aspekter som förmågan att kommunicera effektivt och samarbeta ska inkluderas (Borman et al. 1983; Brief & Motowidlo 1986; Organ 1988; Campbell 1990; Campbell et al. 1996; Organ 1997; Rotundo & Sackett 2002). Ytterligare ett omdiskuterat ämne är hur man kan eftersträva personalsammansättningar som är representativa för ett lands demografi samtidigt som man använder selekteringsmetoder med hög validitet (Barrett & Luecke 2004; Grutter v. Bollinger et al. 2003; Sharf & Jones 1999). En tredje stor fråga är hur man genom selekteringsmetoder eller selekteringsstrategier främjar hög prestation inom arbetslag, och inte bara hög prestation på individuell nivå (Kozlowski & Klein 2000).

Slutsatsen av denna generella genomgång av existerande selekteringsteori är att det, med undantag för den mest övergripande processen för selektering, fortfarande pågår en akademisk diskussion angående de flesta koncept. En utbredd konsensus för hur man ska förhålla sig till selektering och vilka metoder man ska använda verkar inte existera.

## **2.2 Selektion baserad på fysiska attribut**

Som uppmärksammats ovan existerar omfattande teoretisk forskning angående generell selektion. Däremot framstår det akademiska arbetet angående selektion baserat på fysiska attribut mer begränsat. Ryan et al. (1999) har gjort en omfattande empirisk undersökning av de olika selekteringsmetoderna som företag använder sig av. Undersökningen involverade 959 företag i 20 olika länder. En av slutsatserna i undersökningen var att test av fysisk förmåga var en av de testtyper som sällan eller aldrig användes. Ryan fokuserade enbart på företag och involverade inte organisationer som kan antas ha större användningar av fysiska tester som militära eller polisiära organisationer, men den begränsade förekomsten av fysiska tester som Ryan uppmärksammade kan möjligtvis förklara varför det akademiska intresset för att undersöka sådana tester är mer begränsat.

Trots en brist på akademiskt material har vi funnit exempel på akademiska arbeten angående selektion baserat på fysiska attribut. En av de mer relevanta är "Isokai-testet som urvalstest till utbildning i Försvarmaktens nya personalförsörjningssystem" (Larsson et al. 2011). Denna rapport från Karolinska Institutet undersöker främst huruvida Isokai-testet mäter muskelstyrka på ett tillförlitligt sätt. Med andra ord om samma sökande får samma

resultat när de utför testet flera gånger, med samma provledare eller med två olika provledare (Larsson et al. 2011). I rapporten konstateras att så är fallet. Resultatet har en relevans för vår studie i den mening att rapporten bekräftar att Isokai-testet har hög reliabilitet, en faktor som Payne och Harveys ramverk tar upp.

Utifrån vår genomgång av relevant teori lyckades vi endast identifiera två akademiska arbeten som använder ett ramverk för att skapa eller utvärdera fysiska selekteringstest. Artikeln *“Content validity index and intra- and inter-rater reliability of a new muscle strength/endurance test battery for swedish soldiers”* (Larsson et al. 2015) undersöker validiteten hos 30, och reliabiliteten hos fem, olika muskelstyrketest som observerats i litteratur eller vanligen används av nordiska alternativt NATO-länder (Isokai-testet undersöks ej). Denna artikel introducerar ett alternativt sätt att analysera fysiska selekteringstest som avviker från ramverket Payne och Harvey introducerar men som till viss del använder samma teoretiska koncept.

Som tidigare nämnts ämnar dock denna studie utgå ifrån artikeln *“A framework for the design and development of physical employment tests and standards”* skriven av W. Payne och J. Harvey (2010). Artikeln presenterar ett teoretiskt ramverk för hur *“physical employment standards”* och *“physical employment tests”* ska utformas där faktorer som validitet och reliabilitet uppmärksammas tillsammans med andra viktiga faktorer.

Utifrån de två artiklar vi har kunnat identifiera anser vi att artikeln presenterad av Payne och Harvey (2010) utgör det mest omfattande och genomgående försöket att generera ett teoretiskt ramverk angående fysiska selekteringstest. Ramverket använt av Larsson et al. (2015) anser vi inte heller vara lika generaliserbart som Payne och Harveys ramverk eftersom ramverkets syfte endast verkar vara att utvärdera de testen som uppmärksammas i artikeln. Payne och Harvey skapar sitt ramverk med det uttryckliga syftet att det ska kunna användas av andra vid skapandet av alla sorters selekteringstest. Av den anledningen har vi valt att undersöka just denna artikel i vår studie. Artikelns ramverk beskrivs utförligt nedan.

## **2.3 A framework for the design and development of physical employment tests and standards**

### **2.3.1 Artikelöversikt**

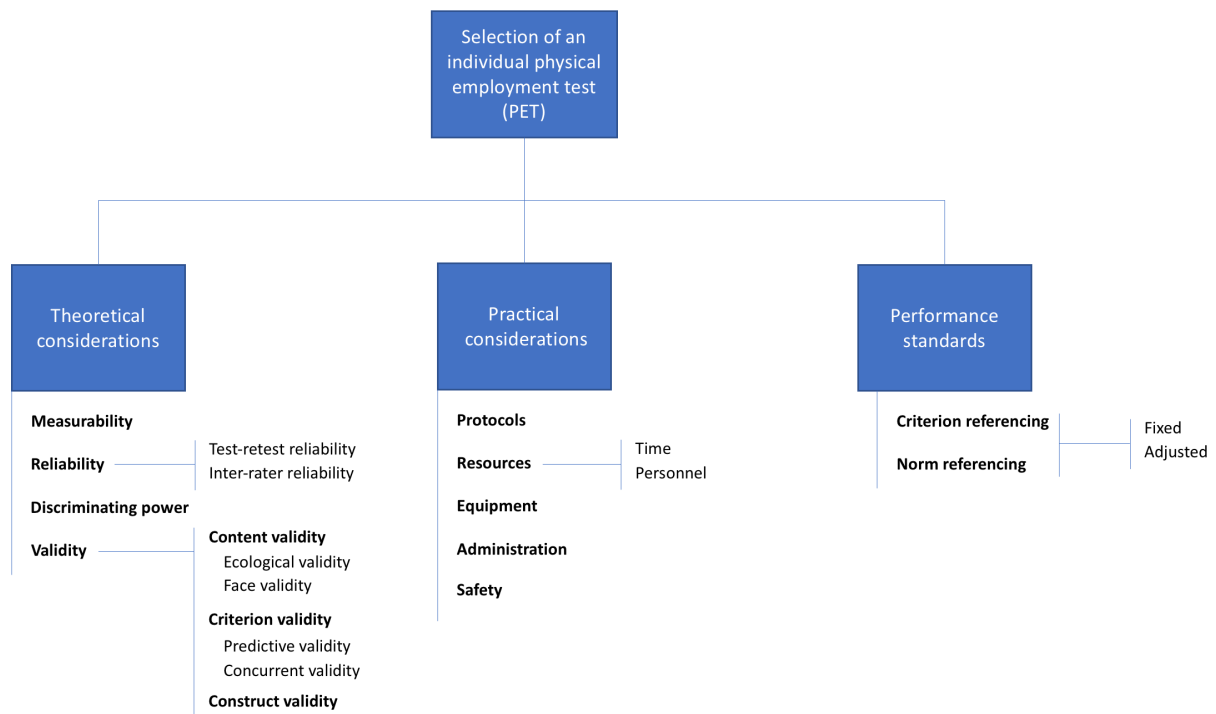
*A framework for the design and development of physical employment tests and standards* (Payne & Harvey 2010) är en artikel skriven med syftet att vägleda i skapandet eller valet av ett, eller flera, selekteringstest baserade på fysiska attribut (PET). Således är artikeln och dess ramverk inte utformat för att utvärdera redan existerande fysiska test.

Artikeln teoretiska bidrag kan summeras i form av en presentation av tre teoretiska konstruktioner angående selekteringstest baserade på fysiska attribut. Den första teoretiska konstruktionen är en omfattande summering av de faktorer som bör tas hänsyn till vid skapandet av ett fysiskt selekteringstest. Den andra teoretiska konstruktionen är en klassificering och indelning av olika fysiska selekteringstest i tre kategorier längs med ett kontinuum. Utifrån denna klassificering redovisar artikeln de fördelar och nackdelar som generellt gäller för tre olika kategorierna av test. Den tredje konstruktionen berör relationen mellan prestationen på testet och prestationen på jobbet. Det vill säga, kopplingen mellan resultat på ett styrketest och efterföljande resultat på jobbuppgifter, vilka begränsningar som finns i den relationen och vilka avvägningar som behöver göras.

Artikeln introducerar ytterligare en del information utöver dessa tre teoretiska konstruktioner som vi ej kommer presentera i detalj. Dels går artikeln igenom behovet av att analysera arbetsuppgifterna för det jobbet man skapar och de fysiska krav dessa arbetsuppgifter ställer på de anställda. Användbarheten av denna *job analysis*-process kommer inte bedömas i denna studie eftersom dess användbarhet redan har etablerats inom selekteringsteori (se avsnitt 2.1). Artikeln tar även upp koncept relevanta vid valet av flera selekteringstest, koncept som inte kommer presenteras i denna text eftersom studien bedömer ramverkets användbarhet vid utvärderingen av ett enskilt test. Slutligen förklarar artikeln de tre teoretiska konstruktionerna mer ingående än vi kommer göra i denna studie, men vi har försökt att göra originalredogörelserna rättvisa givet vårt begränsade utrymme.

### **2.3.2 Summering av viktiga faktorer vid skapandet av ett fysiskt selekteringstest**

Artikeln introducerar en tabell med översikt över "*Key issues in the design of physical employment tests (PET) and physical employment standards (PES)*". Denna tabell summerar alla de aspekter och underliggande faktorer som enligt artikeln bör uppmärksammas och analyseras vid skapandet eller valet av ett eller flera fysiska tester. Eftersom denna studie fokuserar på utvärderingen av ett enskilt fysiskt test summerar vi endast de aspekter och faktorer från denna tabell som är relevanta för ett enskilt test i figur 1 nedan.



Figur 1 - Aspekter och dess underliggande faktorer, introducerade av Payne och Harvey (2010)

I artikeln väljer Payne och Harvey (2010) att sortera in de faktorer som bör tas hänsyn till under tre olika aspekter: *Theoretical considerations*; *Practical considerations* och *Performance standards*.

### 2.3.2.1 Theoretical considerations

Under *theoretical considerations* presenteras fyra olika faktorer. Den första av dessa, *measurability* (mätbarhet) hänvisar till huruvida resultat på ett test är kvantifierbara (på en bedömningsskala eller mer exakt i form av exempelvis prestationstid eller vikt man lyft) eller om resultat är binärt i sin natur och inte går att mäta mer exakt än godkänt/icke-godkänt. Ett test som inte producerar kvantifierbara resultat är inte ett lämpligt PET.

Ett PET vars resultat är binärt har också en begränsad *discriminatory power*. En test med hög *discriminatory power* kan exempelvis dela in testtagare i olika grupper beroende på prestation. Detta är möjligt med test som har kvantifierbara resultat men inte med de test som endast kan producera binära resultat. *Discriminatory power*, som begreppet används i denna artikel, är alltså en positiv egenskap hos ett selekteringstest och ska inte blandas ihop med begreppet diskriminering och ordets användning i dagligt tal.

Ett tests reliabilitet är testets förmåga att producera samma resultat upprepade gånger, när prestationen, eller testtagarens förmåga, är densamma. Reliabilitet delas upp i två faktorer, *test-retest reliability* samt *inter-rater reliability*. *Test-retest reliability* innebär att testet ska ge konsekventa (liknande) resultat om testet genomförs vid två olika tillfällen medan *inter-*

*rater reliability* innebär att testet ska ge konsekventa resultat när testet genomförs två gånger, med hjälp av två olika grupper av testpersonal, vid samma tillfälle.

Ett tests validitet indikerar i vilken utsträckning ett test mäter vad det är designat för att mäta. Det finns flera olika typer av validitet, varav de som Payne och Harvey (2010) anser vara de tre huvudsakliga är *content validity*, *criterion validity* samt *construct validity*. Två typer av validitet som är närbesläktade med *content validity* är *ecological validity* och *face validity*, medan *criterion validity* består av två subtyper, *predictive validity* samt *concurrent validity*.

*Content validity*, när det kommer till PET, innebär att ett test i tillräcklig utsträckning kan representera de fysiska krav som ställs på individen i arbetsrollen. *Ecological validity* relaterar starkt till *content validity* genom att *ecological validity* indikerar i vilken utsträckning testets metoder, material och miljö liknar de som återfinns i arbetssituationen. *Face validity* anger hur uppenbart testets relation till jobbet ifråga är, i synnerhet för de som genomgår testet och inte har någon erfarenhet eller expertkunskap om jobbet.

*Criterion validity* förklarar den statistiska relationen mellan erhållna värden på testet och erhållna värden på ett direkt mått av arbetsprestationen vid senare arbete (*predictive validity*) eller erhållna värden på ett oberoende indirekt mått av arbetsprestation (*concurrent validity*; exempelvis erfaren handledares bedömning av arbetsprestation).

*Construct validity*, i artikelns kontext, hänvisar till ett tests statistiska relation i förhållande till en underliggande fundamental fysiologisk kapacitet, exempelvis aerobisk kapacitet eller maximal helkroppstyrka.

Ett test ska alltid ha antingen *content validity* eller *criterion validity* för att anses vara ett test som mäter det som det är designat att mäta, det vill säga vara tillräckligt validerat. Enbart *construct validity* anses inte tillräckligt för att bevisa en koppling.

### **2.3.2.2 Practical considerations**

*Under practical considerations* uppmärksammar artikeln fem faktorer: *protocols*, *resources*, *equipment*, *administration* samt *safety*. Innebörden av dessa är lättförståeliga och förklaras endast i begränsad omfattning i artikeln. Utöver de teoretiska hänsynstagandena måste testet vara praktiskt genomförbart med hänsyn till den utrustning, tid och personal som testet kräver. Det krävs även att testprotokoll, testinstruktioner och fungerande administration etableras för både de som administrerar testet och de som genomför testet. Slutligen måste testet implementeras på ett säkert sätt, det måste vara säkert att konstruera testet och säkert att administrera och genomföra testet.

### 2.3.2.3 Performance standards

Den sista aspekten att ta hänsyn till enligt artikeln är *performance standards* (även kallat *test standards*), med andra ord de standarder man behöver uppnå på ett test för att anses godkänd, eller för att uppnå någon annan nivå eller kategori. Det finns två sätt att härleda en teststandard. Antingen kan man basera en standard på de faktiska fysiska krav som man etablerat behövs för att genomföra jobbet på ett tillfredsställande sätt (*criterion referencing*), alternativt kan man basera en standard på en viss populations distribution (*norm referencing*). Populationen som refereras till kan exempelvis vara prestationsnivåerna som mäts hos de som nu innehar jobbet ifråga eller prestationsnivåerna hos de kandidater man tidigare testat.

Båda former av standarder kan antingen vara *fixed* eller *adjusted*. *Adjusted standards* innebär att man justerar standarden som krävs för godkänt resultat baserat på individuella karaktärsdrag hos testpersonen eller på grund av avvikelser i miljön där testet genomförs. Exempel på individuella karaktärsdrag är exempelvis kön, ålder eller kroppsstorlek medan exempel på avvikelser i miljön är regn eller värme.

Att korrigera *criterion-referenced standards* baserat på individuella karaktärsdrag kan inte försvaras utifrån motivet att bättre vilja förutse en individs arbetsprestation utan görs i så fall för att uppnå önskvärda sociala eller politiska målsättningar, exempelvis öka andelen kvinnor inom ett yrke (positiv diskriminering). Att korrigera standarden för alla individer vid avvikelser i testmiljö är dock motiverat för att uppnå bättre prediktion. *Fixed criterion-standards* kan vara diskriminerande mot individer om valideringen av testet är otillräcklig. Till exempel är det diskriminerande mot kvinnor att applicera samma teststandarder på kvinnor som på män, om de ursprungligen utvecklades endast för män. Detta gäller även om de efterföljande arbetsuppgifterna är desamma för kvinnor och män.

Det finns två olika sätt att applicera *norm referencing*. Vid *determinative norm-referencing* sätter man godtyckligt en gräns med hänsyn till en populations distribution (*determinative norm-referencing*) utan någon användning av *criterion referencing*. Ett exempel på en godtycklig gräns skulle kunna vara att resultat bättre än genomsnittet krävs för godkänt resultat utan någon vidare hänsyn till om denna nivå krävs för att utföra jobbet eller ej. *Confirmatory norm-referencing* innebär å andra sidan att man använder *norm referencing* som ett komplement till *criterion referencing*. Standarden som skapats utifrån *criterion referencing* ställs i relation till vad som faktiskt presterats tidigare för att avgöra om kravbilden är realistisk. Medan *confirmatory norm-referencing* är användbart och försvarbart är *determinative norm-referencing* arbiträrt och inte försvarbart. Således krävs alltid *criterion referencing* för att etablera försvarbara teststandarder.

### 2.3.3 Kategorisering av fysiska selekteringstest

Den andra stora teoretiska konstruktionen som Payne och Harvey (2010) introducerar är en indelning av alla potentiella fysiska selekteringstest i tre kategorier. De tre kategorierna är: *Generic predictive tests* (GPT), *Task-related predictive tests* (TPT) samt *Task simulation tests* (TST).

GPT mäter en specifik fysisk förmåga. Utförandet av testet har inga direkta likheter med utförandet av någon jobbuppgift men kan användas för att försöka förutspå prestationen i en eller flera jobbuppgifter. TPT baseras inte på en specifik jobbuppgift utan på en mindre resurskrävande jobbrelaterad aktivitet, eller på en aktivitet med egenskaper som relaterar till jobbet på ett mer generellt plan. Ett TST baseras istället direkt på en representativ jobbuppgift.

Dessa tre kategorier är inte diskreta, utan GPT och TST representerar två extremer på ett kontinuum medan TPT kombinerar element från de båda och hamnar i mitten av kontinuumet. Ett test kan således placera sig någonstans emellan alla tre kategorier.

Fördelar och nackdelar med respektive kategori av test summeras i tabell 2 nedan, tabellen är baserad på en tabell presenterad i artikeln, med aspekter bedömda som mindre relevanta för vår applicering utelämnade.

	<b>Generic predictive test</b>	<b>Task-related predictive test</b>	<b>Task simulation test</b>
<b>Etablering av testets reabilitet</b>	Möjligtvis ej nödvändigt om etablerat i akademisk litteratur	Nödvändigt	Nödvändigt
<b>Content validity i relation till en specifik jobbuppgift</b>	Låg	Hög	Väldigt hög
<b>Predictive validity i relation till en specifik jobbuppgift</b>	Okänd	Antagligen hög	Väldigt hög
<b>Predictive validity i relation till andra jobbuppgifter</b>	Okänd, men antagligen rimlig	Okänd	Okänd
<b>Representativt för alla specialisering inom en jobbkategori</b>	Potentiellt	Begränsat	Begränsat
<b>Antal test som behövs för att testa flera specialiseringar</b>	Få	Många	Många
<b>Konsekvenser av att använda test för selektering till flera specialiseringar</b>	Lägre kostnad, relativt enkelt, generiska vanliga test	Potentiellt kostsamt, komplext och omfattande, specialiseringsspecifika test	Potentiellt kostsamt, komplext och omfattande, specialiseringsspecifika test
<b>Applicerbarheten av normativ data</b>	Viss resultatlikhet mellan specialiseringar	Kan vara specialisationsspecifik	Kan vara specialisationsspecifik

Tabell 2: Styrkor och svagheter med respektive kategori av test.

### 2.3.4 Relationen mellan prestation på testet och prestation på jobbet

Relationen mellan testresultat och prestationsnivå gällande specifika arbetsuppgifter mäts med hjälp av statistiska metoder som *Pearson correlation coefficient* samt *linear regression analysis*. Ett stickprov av testkandidater som genomfört både testet och arbetsuppgifterna används för att etablera nivån av *predictive validity*. Prestationsnivåer på test och arbetsuppgifter dikotomiseras vanligtvis, det vill säga delas in i två läger, en grupp värden som bedöms som en godkänd prestation på antingen ett test eller en arbetsuppgift, medan en grupp värden bedöms som en icke-godkänd prestation.

Så länge relationen mellan testresultat och prestationsnivån på en eller flera arbetsuppgifter inte är perfekt (ett visst prestationsvärde på testet ger alltid ett motsvarande prestationsvärde på arbetsuppgiften) kommer det på ett dikotomiserat test förekomma *false negatives* (icke-godkänd prestationsnivå på testet men godkänd nivå på arbetsuppgiften) samt *false positives* (godkänd prestationsnivå på testet men icke-godkänd nivå på arbetsuppgiften). Genom att sänka kravet för godkänt på ett test sänker man mängden *false negatives* men ökar mängden *false positives*. Höjer man kraven resulterar det i motsatt situation. För att sänka både mängden *false negatives* och *false positives* behöver relationen mellan prestationsvärdena på testet och prestationsvärdena på arbetsuppgiften (*predictive validity*) förbättras. Som kan observeras i tabell 2 förbättras nivån av *predictive validity* generellt när testet karaktär rör sig längs kontinuumet från GPT mot TST.

## 2.4 Problematisering av begreppet validitet

Ett centralt begrepp i ramverket ovan, och även i många av de teorier och modeller som presenterades i avsnitt 2.1, är validitet. Således kan det konstateras att validitet och specifikt olika selekteringsmetoders validitet, är ett avgörande begrepp inom selekteringsteori.

Validitet är dock omdiskuterat som begrepp (Newton 2013). Validitet är inte bara relevant vid selektering, utan begreppet appliceras inom många olika akademiska discipliner. Inom psykologin, ett relevant område för selektering, finns det över hundra olika typer av validitet (Newton 2013).

Det första steget i den senare vanliga tendensen att dela upp validitet i olika typer skedde 1954 då *American psychology association* (APA) tillsammans med två andra akademiska organisationer för första gången etablerade en rådande standard för begreppet (APA, AERA & NCMUE 1954). De rekommenderade 19 olika validitetstyper varav fyra kom att få väldigt utbredd användning. Dessa typer av validitet var: *content validity*, *concurrent validity*, *predictive validity* samt *construct validity*.

Samma organisationer kom senare att publicera nya standarder flera gånger: 1966, 1985 samt 1999 (Newton 2013). I den senaste versionen av standarder, publicerade 1999, talar man om bevis baserat på olika fenomen som: *test content, response processes, internal structure, relations to other variables* och *consequences of testing*. Dessa olika bevis sägs tydliggöra olika sidor av validitet, men validitet anses idag vara ett enhetligt begrepp som inte på ett tydligt sätt kan delas upp i olika typer validitet (Newton 2013).

Strax efter att APA publicerade sin första standard för validitetsbegreppet kom en uppsjö akademiker att börja publicera arbeten som introducerade nya typer av validitet (Loevinger 1957; Campbell & Fiske 1959; Cureton 1965). Denna trend fortsatte efter introduktionen av nya standarder (Bemis 1968; Carver 1974; Ebel 1983; Foster & Cone 1995). Faktum är att denna akademiska trend fortsätter med oförminskad styrka än idag, trots att APA förkastat idén med flera typer av validitet (Freebody & Wyatt-Smith 2004; Shaw & Weir 2007; Karelitz et al. 2010; Guion 2011) . Denna ihållande trend är vad som gett upphov till den omfattande mängd validitetstyper som existerar idag. Trenden indikerar också att den eventuella konsensus som APA skapat tillsammans med andra organisationer är i svagaste laget och ofta ifrågasätts inom akademien. Det är fortfarande vanligt att arbeta med begreppet validitet i form av olika typer av validiteter och således finns lite som tyder på att detta teoretiska synsätt generellt anses förlegat.

*A framework for the design and development of physical employment tests and standards*, den artikel som denna studie baseras på, hanterar således validitet på ett traditionellt sätt (Payne & Harvey 2010). Det bör därför uppmärksammas att artikeln hanterar validitetsbegreppet på ett sätt som av APA tillsammans med andra organisationer har bedömt som förlegat. Vi observerar dock att denna hantering av validitet är utbredd inom akademien idag och anser att APA:s bedömning inte är skäl nog att diskvalificera artikeln och dess användning av validitetsbegreppet.

## **3. Metod**

### **3.1. Metodval och forskningsansats**

#### **3.1.1 Val av ansats**

Studien utgick huvudsakligen från en deduktiv ansats eftersom studien hade utgångspunkten i befintlig teori (Bryman & Bell 2013). En extensiv litteraturstudie utfördes tidigt i processen och med förståelsen som bildades skapades en mall med frågor för de kvalitativa intervjuerna för möjligheten till empirisk insamling av bredd och djup. Studien visade sedan på induktiva inslag då den insamlade empirin i samband med ökad förståelse under studiens gång medförde uppkomsten av forskningsfråga 2 och en analys som identifierade två fenomen som bör vara del av det befintliga ramverket (Glaser & Strauss

1967). Således blev studiens forskningsansats abduktiv på grund av kombinationen av deduktiv och induktiv metod. En abduktiv metod är ofta fördelaktig då deduktiv metod kan kritiseras för brist i klarhet gällande vilken teori som initialt ska väljas medan induktiv metod kan kritiseras då oavsett mängd empiri behöver denna inte nödvändigtvis medföra teoribildning (Saunders, Lewis & Thornhill 2012).

### **3.1.2 Forskningsstrategi**

Vid val av studie föreföll kvalitativa intervjuer lämpligt eftersom det som efterfrågades var respondenternas kunskaper och erfarenheter gällande Isokai-testet. Med hjälp av kvalitativa intervjuer ansåg vi att vi på bästa sätt skulle kunna utreda forskningsfrågan genom att låta respondenterna beskriva sina erfarenheter med Isokai-testet och få insikt i deras tid i Försvarmakten och därigenom kunna bedöma användbarheten av, och eventuella brister i, Payne och Harveys ramverk. Vidare bedömdes kvalitativ metod lämplig eftersom vår uppfattning var att Försvarmaktens antagningstester diskuteras och kritiseras flitigt av de anställda och att vi därav gjorde antagandet att det skulle kunna förekomma starka åsikter, och eventuellt element av känslig karaktär.

### **3.1.3 Forskningsdesign**

Studien genomfördes i form av en fallstudie för ett specifikt fall rörande en organisation med ett mål om 15-20 intervjupersoner från organisationen (Bryman & Bell 2013). Enligt Yin (2003) är fallstudier lämpliga då man ämnar undersöka och få en djupgående förståelse för ett fenomen i dess naturliga miljö. Kritik kan föras mot forskningsdesignen då den endast hanterade en organisation och därav undgick möjligheten till jämförelse mellan olika fall som i multipla fallstudier. Ett alternativ skulle kunna ha varit att även studera rekryteringen och selekteringen till Polishögskolan som också använder Isokai-testet för att testa sökande kandidater. Vi fann dock att ett bredare tillvägagångssätt med fler fall hade äventyrat studiens förmåga att leverera tydliga slutsatser, i synnerhet med hänsyn till det begränsade utrymmet. Fördelen med ett enskilt fall i jämförelse med en flerfallstudie är att man kan ägna mer uppmärksamhet åt den specifika kontexten i fallet snarare än att fokusera på att jämföra flera fall (Dyer & Wilkins 1991). Trots att vi tog valet att inte göra en flerfallstudie kan man resonera för att resultatet kan komma få en viss grad av generaliserbarhet för andra fall, som exemplet med Polishögskolan ovan eftersom samma test används. Dock hävdar Lee med flera (2007) att just partikularisering snarare än generalisering är fallstudiernas främsta styrka varför målet i studien bör vara att utveckla en djupgående förståelse för komplexiteten i det unika fallet. Därför anser vi att denna studies eventuella generaliserbarhet bör beaktas med försiktighet.

## 3.2 Urval

### 3.2.1 Val av organisation

Fallföretaget som valdes för studien är den svenska försvarsmakten. Valet att inrikta studien på Försvarsmakten gjordes inledningsvis på grund av vårt personliga intresse för organisationen. Vi resonerade även att en stor offentlig organisation skulle kunna ge oss goda möjligheter att få kontakt med anställda och andra delaktiga vilket skulle minska risken för bristande tillgänglighet gällande information och intervjupersoner.

### 3.2.2 Val av intervjupersoner

Totalt genomfördes 15 djupintervjuer av personer som är eller har varit aktiva i Försvarsmakten. Utav de 15 intervjuade identifierade sig åtta som män och sju som kvinnor. Tabell 3 nedan listar samtliga intervjuades ålder, befattning, tid i Försvarsmakten samt om de är aktiva idag. Kolumnen längst till vänster visar intervjupersonernas benämningar i studiens empiri.

Benämning	Ålder	Befattning	Tid i FM	Aktiv i FM idag	Intervjudatum	Typ	Intervjuad av
Respondent 1	24	Tolk	3 år	Nej	2017-04-01	Närvarande	Båda
Respondent 2	27	Ställföreträdande gruppchef	7 år	Ja	2017-04-01	Närvarande	Båda
Respondent 3	24	Endast GMU	3 mån	Nej	2017-04-01	Närvarande	Båda
Respondent 4	23	Endast GMU	3 mån	Nej	2017-04-01	Närvarande	Båda
Respondent 5	24	Kadett - Officersprogrammet	3 år	Ja	2017-04-02	Telefon	Båda
Respondent 6	22	Skyttesoldat	3 år	Ja	2017-04-02	Telefon	Båda
Respondent 7	22	Kadett - Officersprogrammet	2 år	Ja	2017-04-03	Telefon	Båda
Respondent 8	23	Stabsassistent	3 år	Ja	2017-04-06	Telefon	Båda
Respondent 9	24	Ledningssoldat - Rekryttingsprocesser	2 år	Ja	2017-04-05	Telefon	Båda
Respondent 10	26	Reservofficer	4 år	Ja	2017-04-05	Närvarande	André
Respondent 11	28	Plutonchef - Sjukvårdstjänst	9 år	Ja	2017-04-06	Närvarande	André
Respondent 12	62	Plutonchef	14 år	Nej	2017-04-07	Telefon	Båda
Respondent 13	21	GMU samt FOK	6 mån	Nej	2017-04-11	Telefon	André
Respondent 14	21	Endast GMU	3 mån	Nej	2017-04-24	Telefon	André
Respondent 15	47	Akademisk expert inom militär rekrytering	5 år	Nej	2017-04-26	Närvarande	Båda

Tabell 3 - Information angående studiens intervjuer och intervjupersoner

Intervjuerna genomfördes med personer som är eller har varit aktiva i den militära delen av Försvarsmakten. Civilanställd personal intervjuades inte eftersom dessa inte genomfört mönstring och därav inte heller Isokai-testet. Initialt kontaktade vi Försvarsmakten via deras officiella kanaler genom de olika regementenas kontaktpersoner. De officiella kanalerna besvarade inte våra förfrågningar i tid nog vilket ledde oss att ta kontakt med bekanta som är eller varit personal i Försvarsmakten. Denna metod visade sig effektiv då de bekanta sammankopplade oss med deras kollegor eller vänner som gärna ville ställa upp. Kritik mot detta tillvägagångssätt är att det kan te sig vara ett bekvämlighetsurval. Detta eftersom

tidsbristen och svårigheterna med officiella kanaler medförde att vi fick tillgång till, och valde intervjupersoner utifrån vad som var smidigt och effektivt vid tidpunkten.

Gällande avgränsningen för intervjupersonernas befattningar, antal år i Försvarsmakten, kön och ålder valde vi att eftersträva en bredd i samtliga kategorier för att öka möjligheterna till varierande svar och erfarenheter i intervjuerna och därmed ett starkt empiriskt underlag. Empirisk mättnad uppnåddes till stor del gällande den stora massan av intervjuerna eftersom liknande erfarenheter och åsikter återkom. Intervjuerna med respondenterna 11 och 15, som besatt större kunskap kring rekrytering och selektering, påvisade andra insikter och därmed annan empiri än vid de andra intervjuerna. Att dessa intervjuer bara var två stycken är en svaghet rörande den empiriska mättnaden. Något som bör tilläggas är att några av de intervjuade genomgick antagningsprövningen för lång tid sedan och hade därför vissa svårigheter med att minnas vissa aspekter eller detaljer av Isokai-testet. Detta har eventuella implikationer för empirins trovärdighet.

### **3.3 Kvalitativ undersökning**

#### **3.3.1 Semistrukturerade djupintervjuer**

Vi valde att använda oss av semistrukturerade djupintervjuer. Inför intervjuerna skapades en intervjuguide med frågor att ställa till intervjupersonerna (se bilaga 1). Guiden utformades efter vi bildat en förståelse i relevant teori för att kunna ställa frågor utifrån många vinklar och perspektiv för att öka möjligheterna till intressant empiri och en stark teoretisk koppling. Introduktionen av guiden utformades narrativt för att få intervjupersonen att känna sig bekväm och kunna påbörja diskussion kring Isokai-testet utifrån ett bredare perspektiv.

Kritik mot intervjuguiden finns i att den kan anses vara utformad som ett frågeformulär att använda vid strukturerade intervjuer och kvantitativ forskning snarare än kvalitativ (Bryman & Bell 2013). Detta kan verka motstridigt med studiens syfte och metod och denna motstridighet var inte planen. Dock var alltid målet att använda frågorna i guiden som en mall under intervjuerna för de teman vi ville beröra vilket sedan fungerade väl när intervjuerna genomfördes. Under intervjuerna tilläts processen vara flexibel och intervjupersonerna fick frihet att utforma svaren på sina egna sätt, frågorna ställdes inte alltid i samma ordning och vi ställde ofta följdfrågor som anknöt till något intervjupersonen sagt (Bryman & Bell 2013). Intervjuerna blev därför semistrukturerade enligt plan trots motstridigheten i utformandet av guiden. Ytterligare kritik mot intervjuguiden är att frågorna som de är formulerade till viss mån är ledande och söker efter vissa svar eftersom de skapades med Payne och Harveys ramverk färskt i minnet.

### **3.3.2 Genomförande av intervjuer**

Utav samtliga 15 intervjuer utfördes sju på plats med personen och åtta över telefon. Vid fyra av intervjuerna kunde bara en av oss närvara på grund av tidsbrist och svårigheter med att hitta andra tider med intervjupersonerna. Att åtta av intervjuerna genomfördes över telefon var inte fördelaktigt med tanke på att vissa element av kommunikation elimineras. Över telefon saknas ögonkontakt, kroppsspråk och faktorer som ökar bekvämligheten och därmed äventyras överföringen av information. Att en stor del av intervjuerna utfördes över telefon förklaras dels av tidsbristen samt att det hade varit praktiskt ogenomförbart att ta sig till alla intervjupersoner eftersom de flesta intervjupersoner befann sig på annan ort. Det bör tilläggas att vi trots det inte anser att intervjuerna över telefon var mindre givande än de fysiska i termer av djup i diskussionerna och utbyte av information.

### **3.3.3 Anonymisering**

Vid inledandet av samtliga intervjuer informerade vi intervjupersonerna om deras anonymitet i uppsatsen och att det enda vi skulle kunna komma att nämna var de karaktäristika som återfinns i tabell 3. Vi uppfattade att de intervjuade kände sig bekväma under intervjuerna och ingen uttalade någon rädsla för att det de sa skulle kunna härledas till dem. Vi ansåg deras medvetenhet kring anonymiseringen medförde en förmåga att tala fritt och bidrog till ett rättvisande empiriskt material. I empirin används "hen" som personligt pronomen för att ytterligare säkerställa att personerna anonymiseras.

### **3.3.4 Metod vid bearbetning av data**

Från att de första fem intervjuerna genomförts påbörjades transkriberingen gradvis. Utav alla 15 intervjuer har tio transkriberats i sin helhet medan fem enbart delvis på grund av bristande tid. I de delvis transkriberade har de delar vi bedömt vara mest givande och relevanta för studiens syfte transkriberats. Vid bearbetandet av empirin har vi kodifierat utifrån den metod som Strauss och Corbin (1990) beskriver som öppen kodning där man bryter ner, studerar, konceptualiserar och kategoriserar data. I empirin presenteras de trender vi kunnat observera utifrån de diskussioner som återkommit och andra observationer som bedömts som mer intressanta i intervjuerna.

## **3.4 Metoddiskussion**

I detta avsnitt används termerna validitet och reliabilitet för att diskutera studiens metod. Användningen av termerna i denna diskussion är helt frikopplad från användningen av samma termer i resterande delar av studien. Bryman och Bell (2013) tar upp LeCompte och Goetz (1982) fyra huvudsakliga begrepp för att bedöma tillförlitligheten i en studie, extern och intern reliabilitet och validitet.

Extern reliabilitet beskrivs av LeCompte och Goetz (1982) som bedömningen av huruvida en studie kan replikeras. I kvalitativ forskning är det oftast svårt att uppfylla då studien till stor del påverkas av den sociala miljön. Replikerbarheten i denna studie kan te sig problematisk då analysen är beroende av just de personer vi intervjuat och ett annat urval skulle möjligen kunna ge andra resultat.

Den interna reliabiliteten bedöms genom att de som genomför en studie tolkar det de ser och hör med överensstämmelse (LeCompte & Goetz 1982). Tolkning av individuell prägel kan rimligtvis förekomma i studien, dock har vi i regel under diskussioner tolkat och uppfattat i linje med varandra.

Extern validitet berör till vilken grad studiens resultat kan generaliseras till andra sociala miljöer eller situationer. Enligt LeCompte och Goetz (1982) är graden av generalisering svag för kvalitativ forskning på grund av begränsade urval i fallstudier. Detta är fallet i denna studie och visar på en svaghet eftersom urvalet och den sociala miljön är unik och därmed bör generaliserbarheten beaktas med försiktighet.

Studiens interna validitet bedöms enligt LeCompte och Goetz (1982) utifrån överensstämmelsen mellan forskarnas observationer och senare utvecklade idéer och hävdar att det ofta är en styrka i kvalitativ forskning på grund av den stora närvaron under lång tid i sociala situationer. I denna studie intervjuades 15 personer från samma organisation med djupintervjuer varför en god överensstämmelse mellan begrepp och observationer kan hävdas finnas.

## **4. Empiri**

### **4.1 Diskussionstrender**

I avsnitten nedan presenteras empirin från intervjuerna huvudsakligen utifrån fem observerade diskussionstrender. Kategorierna har valts ut då de har uppmärksammats av ett antal intervjupersoner som intressanta och viktiga diskussioner att ha angående Isokai-testet. Avsnitten 4.1.6 och 4.1.7 behandlar inte diskussionstrender, på så sätt att majoriteten berörde det, utan avsnitten presenterar uttalanden som visade sig vara av stort värde för efterkommande analys.

#### **4.1.1 Kopplingen mellan Isokai-testet och de efterföljande fysiska arbetsuppgifterna**

Samtliga intervjupersoner har frågats angående deras åsikt kring Isokai-testets relevans för vad den sökande senare gör i Försvarmakten. Majoriteten av de intervjuade har negativa åsikter om Isokai-testets koppling till de arbetsuppgifter som senare utförs medan några få tvärtom anser att testet var en bra indikation av de prestationer som senare krävs.

På frågan om den generella åsikten kring Isokai-testets säger respondent 1 att de har diskuterat det med kollegor och att det oavsett befattning: *"...inte det säger så mycket om någonting egentligen"*. Detta indikerar att en diskussion existerar bland personalen som målar Isokai-testet som ett problem. Vid den mer specifika frågan huruvida testet är relevant för det personen gör inom Försvarsmakten svarar respondent 1: *"Inte alls, jag gick tolkskolan"* och fortsätter förklara hur hans vidareutbildning framförallt består av att plugga glosor inomhus och hur lite de aktiverar sig fysiskt under utbildningen. Hen fortsätter med att säga: *"...eller det är såklart bra när man ska bära något eller lyfta upp en bår, men det är det enda jag kan komma på."* På liknande sätt beskriver respondent 9 att hen fått höra att Isokai-testet kan likställas med att lyfta upp en kollega i militär ambulans men att: *"Jag har aldrig gjort det lyftet i tjänst under mina 1,5 år så den anledningen funkar inte, nej jag ser ingen relevans i det"*. Respondent 2 säger: *"Generellt tycker jag att det är ett ganska dåligt test för att, just den rörelsen, att man lyfter någonting ända upp till hakan i det greppet, det gör man ju inte annars."* och respondent 8 belyser på liknande sätt att det typet av lyft så högt upp under hakan sällan förekommer och därav är missvisande för personens styrka.

Respondent 3 svarar mer positivt gällande testets koppling till praktiken och säger: *"Alltså, på ett sätt ja, eftersom det är många, man gör ju många lyft på det sättet, att man lyfter mycket från marken, tunga ryggor och så sätter man dom över huvudet, då ska man ju ha en bra benstyrka, bra ryggstyrka och bra axelstyrka. Så på så sätt är det ju bra. Jag antar att Isokai är relevant men det kanske skulle behövas någon form av komplettering i styrka."*

Respondent 11 som jobbar med rekrytering och utbildning, berättar hur hen skickar alla sina kvinnliga soldater på uppföljningstester i Isokai-maskinen och hur många av de drar sämre resultat efter avklarad GMU.

Det bör tilläggas att vi ställer frågan om de intervjuade hade förtroende för testets legitimitet och koppling till senare arbetsuppgifter när de först kom på plats för att göra antagningstestet. Respondent 8 berättar: *"Nej, alltså tyckte inte att jag fick någon bra instruktion på vad det syftade till och på vad jag skulle göra på riktigt, nja såg lite fjuttigt ut när man kom in"* medan respondent 4 säger: *"Ja, jag tror det. Ja, mitt intryck var väl att det mäter styrkan i hela kroppen och på så sätt är det väl relevant."*

#### **4.1.2 Isokai-testets mätbarhet och reliabilitet**

Bland alla intervjuade individer finns en stark konsensus kring åsikten att testet har en hög mätbarhet och reliabilitet. Flera uppmärksammar att styrkeprestationen vid testtillfället registreras med noggranna och exakta värden. Många lägger även fokus på att testet sker i en kontrollerad miljö, i en konfigurerad dedikerad testmaskin tillsammans med utbildad personal. Alla dessa aspekter anses höja reliabiliteten och mätbarheten.

Endast en individ, respondent 8, anser att testet har låg reliabilitet: *“Nej, för man får ju så olika instruktioner, en i min grupp behövde bara dra till bröstvårtorna medan jag behövde dra till hakan, och det är ju en betydande del, det är där det är som tyngst, så det är inte rättvist, man får olika resultat beroende på instruktioner [...]”*.

#### **4.1.3 Diskriminering genom testdesign och kravnivå**

I intervjuerna diskuteras även huruvida Isokai-testet är diskriminerande med hänsyn till att män och kvinnor utför exakt samma test och har samma kravnivåer för bli antagna till Försvarsmakten och organisationens olika tjänster (se frågorna 13 och 14 i bilaga 1).

Både de män och kvinnor som har intervjuats anser att Isokai-testet möjligtvis kan vara diskriminerande mot kvinnor. De uppmärksammar att testet antagligen har utformats med hänsyn till den manliga kroppen eftersom det introducerades under mitten av 1990-talet då nästan enbart män genomförde mönstringen. I en relaterad diskussion poängteras att kvinnor i genomsnitt har mycket svårare att genomföra den senare delen av det långa draget, när stängen ska föras från navelhöjd upp till hakan.

Alla förutom en intervjuperson anser dock att införandet av olika kravnivåer för män och kvinnor vid antagning till en och samma tjänst är fel väg att gå, samma negativa inställning observeras även för idén om att införa två separata test. Det dominerande åsikten är att om Isokai-testet kan bevisas vara diskriminerande bör det ersättas av ett nytt, mindre diskriminerande test. Gällande kravnivåer anser en majoritet att man, om det är nödvändigt att sänka kraven för att få in fler kvinnor, bör kraven sänkas för både män och kvinnor.

#### **4.1.4 Isokai-testets praktiska genomförbarhet**

De allra flesta intervjupersoner uppmärksammar, antingen på eget initiativ eller efter en relaterat fråga, att en av Isokai-testets stora fördelar är hur praktiskt genomförbart det här. De uppmärksammar bland annat att testet går väldigt snabbt att genomföra. Uppgifterna om hur långt tid det tar varierar mellan en till fem minuter. En annan aspekt som många intervjupersoner nämner är att testet enbart kräver att en observatör/instruktör närvarar. Denna individ behöver inte vara läkare eller specialistutbildad utan är oftast en sjuksköterska. En tredje observation är att Isokai-maskinen är direkt kopplad till en dator vilket möjliggör smidig och digital administrering av resultat och statistik. En typisk beskrivning av Isokai-testet ges av respondent 4: *“Jag fick uppfattningen att det är en relativt enkel maskin och dator att hantera. Att dom bara står där och mäter ens resultat och så gör vi arbete eller vad man ska säga.”* Hen fortsätter: *“Jag var ju inne där i knappt fem minuter, dom förklarade på 30 sekunder, jag gjorde två lyft, dom gav mitt resultat och sen fick jag gå. Det var väldigt effektivt, absolut.”*

Respondent 15 håller med om Isokai-testets praktiska genomförbarhet och effektivitet men ifrågasätter vikten av denna fördel i dagens svenska försvarsmakt: *“Ja, sen kan man väl säga att det utformades då i en era då man testade, vad det var, 25 000 personer om året i alla fall, [...] sen vet jag inte hur många man ska faktiskt fysiskt mönstra om det är 10 000 eller något sådant. Att man eventuellt i alla fall kan tänka sig en liten dyrare lösning med tanke på att det är färre, en dyrare och en mera prick-på lösning i och med att det är färre [...]”*.

Respondent 11, en plutonchef som länge arbetat med rekrytering, framför tidigt i intervjun att hen är medveten om kritik mot testet men att: *“det handlar om vad som är praktiskt genomförbart och vad som är tillräckligt bra för en så stor mängd som ska testas [...]”*.

Många av diskussionerna kring selekteringstests praktiska genomförbarhet centrerade kring de monetära kostnaderna för ett test, vilket kan urskiljas i respondent 15s citat ovan. Ofta antas en positiv relation mellan testets kostnad och förmåga att mäta attribut relevanta för senare arbetsuppgifter, något som citaten ovan från både respondent 15 och 11 indikerar.

#### **4.1.5 Diskussionen kring alternativa test**

Under intervjuerna efterfrågar vi en diskussion kring alternativa sätt att testa sökanden istället för Isokai-testet. Möjliga alternativ till Isokai-testet diskuteras i ljuset av de andra kategorierna vi presenterar och diskussionerna belyser fördelar och nackdelar hos Isokai-testet utifrån andra perspektiv. Så gott som samtliga diskuterar den praktiska genomförbarheten av olika sorters test, ett exempel är respondent 9 som berättar: *“Det skulle i så fall vara något som är mer relaterat till något som är längre och mer pågående, det skulle ju dock kräva mer resurser och tid, för mycket i Försvarsmakten är ju ett långvarigt arbete, även när man är trött ska man vara arbetsduglig så ska man orka, ha envisa muskler [...]”*.

Respondent 15 som har internationell erfarenhet av antagningstester till försvarsmakter använder Kanadas försvarsmakt som exempel vid utformande av tester. Hen inleder diskussionen med: *“Men sen är det ju också det där, vad är det egentligen man mäter och vad som är relevant styrka då?”* och berättar hur Kanadensarna tidigare använt sig av armhävningar, situps och liknande enkla klassiska övningar för att mäta styrkan men att de kom fram till att det dessa tester inte representerade krigets krav. Istället har kanadensarna nu utformat mer praktiska tester, bära sandsäckar, dra en 100 kg tung docka 30 meter och liknande för att testa för faktiska moment i strid. Hen fortsätter med: *“[...] och har då ändå gemensamma krav, så man ser till att det inte finns några sådana tekniska aspekter eller genusaspekter som förfördelar vissa och man är noga med att det ska va då, kopplat till krigets krav och inte något annat.”*

Ett flertal av de intervjuade anser att det vore mer representativt att anpassa Isokai-testet till att mäta prestation över flera repetitioner. Respondent 15 förklarar hur: *“[...] det här är*

ju återigen kopplat till den faktiska verksamheten. Ibland så är det den extrema maxlyftet som krävs. Oftast så är det endurance liksom som är det egentliga kravet.” och att det hade varit bra att testa både maximal styrka och uthållig styrka i den befintliga maskinen och gällande det praktiska anser hen att: “[...]då lägger vi ändå bara till 2 minuter till.”.

#### 4.1.6 Testandet av nuvarande styrka eller potential styrka

Två intervjupersoner fokuserar mycket av sin diskussion på att Isokai-testet mäter nuvarande och inte potential styrka och att konsekvenserna av detta är viktiga att uppmärksamma. Respondent 11 berättar hur resultat på Isokai-testet inte är av relevans senare under utbildningen: *“Det [testet] är användbart men på ett sätt så måste man ha med, det här är ett test som används för testa för att man ska kunna komma in i Försvarmakten, sen när man väl är inne i Försvarmakten, jag som utbildningsplutonchef går ju inte in och tittar på vad mina soldater har fått på Isokai, det skiter jag högaktningfullt i, nu är de hos mig punkt slut.”*. När vi sedan frågar vad hens åsikt var gällande att testet ska motsvara de fysiska krav som sedan krävs under utbildning och vidare tjänst fortsätter hen och säger: *“Nej det behöver nödvändigtvis inte vara så, styrkan är det lättaste att ändra på, styrka och kondition kan man alltid träna upp. Vi kan alltid lära någon att bli stark, det är det minsta problemet.”*. Hen berättar vidare att på tre månader kan hen bygga upp styrkan hos vem som helst.

Respondent 15 belyser hur det är problematiskt med ett så pass smalt och begränsat test som Isokai-testet inför anställning och att det utgör en tröskel: *“Man kan ju säga att det man mäter är ju, nuvarande styrka helt enkelt, inte potential. Så en person som aldrig har tränat och som drar en sju, kanske har otroligt mycket större potential än en idrottsman som går dit och drar ett 8 eller en 9, som redan är i sin fulla prakt på något sätt. Medan om man då tillåter den här som aldrig har tränat och som visade sig ha talang, att träna ett år under grundutbildning, så kanske den personen har en betydligt större fysisk förmåga när den väl kliver in i yrket som färdigutbildad.”* Hen tillägger även att: *“Så på så sätt så blir det [Isokai-testet] ju diskriminerande och kan få utslag inte bara mot dom som, utan det blir vissa grupper då som hamnar, som får för dåliga resultat i förhållande till deras egentliga styrka.”*. Hen anser vidare att denna företeelse kan leda till att man går miste om bra personal.

#### 4.1.7 Indirekta kostnader

Endast en respondent uppmärksammar indirekta kostnader relaterade till selekteringstest i allmänhet och Isokai-testet i synnerhet. Detta är respondent 15 med omfattande expertis och erfarenhet av rekrytering inom försvarsmakter. Som nämnts i kategori fyra uppmärksammar respondent 15 likt de flesta andra att Isokai-testet har låga direkta kostnader. Hen varnar dock för att utvärdera test enbart baserat på direkta kostnader och poängterar att det var minst lika viktigt att titta på indirekta kostnader associerade med ett

visst rekryteringstest: *“Just det här har de [den kanadensiska militären] prioriterat, att få rätt tester och det är ju lätt att titta på kostnaden för ett test liksom och spara pengar där. Men om det innebär att 20 % droppar av, liksom dom nästkommande åren så är det en betydligt större kostnad än att dubbla kostnaden på testet. För då har man investerat ett års utbildning i en person som sen får knäproblem eller vad det nu är.”*

Hen menar att om man tar hänsyn till båda typer av kostnader är det möjligt att den totala kostnaden blir lägre ifall man investerar ordentligt i mer direkt kostsamma test, som samtidigt ger en bättre indikation på en sökandes prestationsnivå och lämplighet för Försvarmakten. Anledningen till detta är att de aspekter som ett bristande test med låga direkta kostnader missar, vilket leder till mer skador och mer avhopp senare: *“[...] och tar man med sådana [indirekta] kostnader, det är ju klart då kan man eventuellt investera betydligt mer i själva testet då, om man kan skära kostnader längre fram.”* Hen menar att ett annat möjligt utfall av att investera mer pengar och resurser i de initiala styrketestet/n är att behålla samma kostnadsnivå men få mer lämpade individer till de olika befattningarna.

Andra respondenter hänvisar inte till indirekta kostnader som ett samlat begrepp eller koncept som man måste ta hänsyn till vid utformningen av ett test men vissa av dessa indirekta kostnader uppmärksammas i andra sammanhang. Många respondenter talar om stora problem med skador under GMU och att Isokai-testet uppenbarligen inte identifierar dessa. Respondent 11 tar också upp respondent 15s andra exempel på indirekta kostnader, avhopp: *“[...] [jag] har utbildat tre kvinnliga piloter, ingen av de jobbar med det idag för de klarar det inte rent fysiskt.”*

#### **4.1.8 Övrig empiri av intresse för analysen**

Fråga 8 i bilaga 1 efterfrågar de intervjuades uppfattning och kunskap kring huruvida Isokai-testet och dess gränsvärden är utformade för att objektivt tillsätta personal utefter objektiva krav eller för att ta in personer som bidrar till homogeniteten i den befintliga gruppen. Majoriteten anser att testet och gränsvärdena är objektivt utformade. Ett exempel är både respondent 15 och 11 som nämner “krigets krav” och där respondent 11 berättar: *“Man brukar prata om krigets krav, en individ får se ut hur den vill men när det väl är krig, då är det krig. Kriget kommer aldrig anpassa sig efter invånarna eller hur någon är byggd.”*

Respondent 4 har ett liknande resonemang men säger även att: *“[...] då är det ju viktigt att dom [kraven] ska korrelera med just den bilden som dom har av att en soldat ska se ut.”* och beskriver vidare att det möjligen kan finnas ett tänk i Försvarmakten att man vill behålla homogeniteten i gruppen.

Ytterligare intressanta diskussioner kommer från fråga 7 från bilaga 1. Här diskuteras resultaten på Isokai-testet i relation till det generella slutgiltiga syftet med antagningsprövningen, att rekrytera individer som bidrar till Försvarmakten. De flesta

intervjupersoner har åsikten att ett väldigt lågt resultat (1-2) på Isokai-testet talade för att en individ skulle ha svårt att bidra till Försvarsmakten som anställd. Skillnader i resultat över denna låga nivå anses dock inte kunna säga mycket om den sökandes senare prestation i Försvarsmakten. Två individer anser att resultat på Isokai-testet inte säger något alls om en individs senare arbetsprestation. Denna diskussion relaterar i stor mån till den tidigare presenterade diskussionen om hur föränderligt styrka som attribut är (se avsnitt 4.1.6). Indikativt på denna koppling är svaret respondent 1 ger på huruvida det finns en relation mellan resultatet på Isokai-testet och individens efterföljande bidrag till Försvarsmakten: "Nej, absolut inte. [...] just därför att styrka är ju något man tränar upp, så det är, fokuset ligger på under utbildningen nästan att träna upp."

## 5. Analys och diskussion

### 5.1 Användbarheten hos de olika delarna i Payne och Harveys ramverk

Alla delar som introduceras i det teoretiska ramverket från Payne och Harveys artikel: *A framework for the design and development of physical employment tests and standards* (2010) nämns och diskuteras av intervjupersoner vid minst ett tillfälle (utan att artikeln, dess faktorer och de teoretiska begreppen har refererats till explicit). Ingen intervjuperson anser att frågor eller ämnen som relaterar till artikelns faktorer är irrelevanta för utvärderingen av Isokai-testet, även om de inte alltid kunnat svara med säkerhet i alla diskussioner.

Det kan konstateras att validitetsfaktorerna visar sig användbara vid utvärdering av Isokai-testet. *Content validity* berörs i alla intervjuer när kopplingen mellan testet och de senare utförda aktiviteterna diskuteras. Exempel finns i empiriavsnitt 4.1.1 där respondenter har erfarenheter som kan beskriva både låg och hög *content validity* i testet. Respondent 15 talar till exempel om "krigets krav" i 4.1.5 vilket är starkt kopplat till *content validity*. Diskussioner kopplat till *ecological validity* återfinns på flera ställen i empirin eftersom det diskuteras mycket om huruvida Isokai-lyftet återfinns i praktiken under tjänst. *Face validity* diskuteras av majoriteten, exempel finns i 4.1.1 då vi efterfrågar specifikt huruvida de intervjuade anser att testet verkade legitimt och representativt för verkligheten vid tidpunkten då de kom i kontakt med Isokai-testet för första gången. *Criterion validity* bör prövas med statistiska metoder men genom fråga 7 i bilaga 1 försöker vi möjliggöra en diskussion kring Isokai-testets *criterion validity* (se avsnitt 4.1.8 i empirin). Alla intervjupersoner för en diskussion kring faktorn och den framstår således som användbar. *Construct validity* som faktor identifieras i flera intervjuer eftersom flera antingen spekulerar, eller som respondent 11 gör i avsnitt 4.1.6, konstaterar att Isokai-testets syfte är att ge en indikation på individens fundamentala styrka snarare än att mäta den sökandes förmåga att utföra arbetsuppgifter.

Faktorerna *measurability* och *reliability* är användbara vid diskussion kring Isokai-testet vilket visas i empirin avsnitt 4.1.2 där de intervjuade oftast tar ställning till dessa. Citatet från respondent 8 gällande olika instruktioners påverkan bör beaktas med försiktighet då det är oklart huruvida instruktioner i samband med testet bör användas för att kritisera testets mätbarhet och reliabilitet då termerna snarare rör testet ur ett tekniskt perspektiv.

*Practical considerations* diskuteras flitigt under merparten av intervjuerna vilket ses i 4.1.4 i empirin. Samtliga fem faktorer diskuteras, främst med fokus på resurserna för utförandet av just Isokai-testet men även vid hypotetiska alternativa testmetoder vilket kan utläsas ur 4.1.5.

Gällande aspekten *performance standards* kan vi se tendenser till både *criterion referencing* då respondent 15 talar om "krigets krav" i 4.1.5 men även *norm referencing*, då respondent 4 beskriver i 4.1.8 hur det kan finnas ett tänk gällande att hålla Försvarsmaktens population homogen.

Isokai-testet, liksom alla andra tester i antagningsprocessen, är samma oavsett kön och testets *performance standards* är *fixed* vilket även kan försvaras utifrån Payne och Harveys ramverk. När vi ställer frågan om sänkta kravnivåer på Isokai-testet för kvinnor, det vill säga att applicera *adjusted performance standards*, svarar majoriteten att det inte vore en bra idé (se avsnitt 4.1.3).

Det är relativt tydligt hur Isokai-testet samt de hypotetiska alternativa test som diskuteras går att placera in under ramverkets tre kategorier av fysiska selekteringstester: GPTs, TPTs och TSTs. Isokai är rimligen ett GPT på grund av det enkla effektiva utförandet och, enligt de flesta intervjupersonerna, relativt låga kopplingen till senare aktiviteter. Respondent 15s utläggning om kanadensarnas tester i 4.1.5 är ett exempel på ett TPT eller TST eftersom dessa är utformade för att representera "faktiska moment i strid". Således verkar artikelns andra teoretiska konstruktion användbar vid utvärderingen av Isokai-testet.

Det verkar även som om den tredje teoretiska konstruktionen som introduceras av artikeln är användbar. Den tredje konstruktionen berör *predictive validity* samt *false positives* och *false negatives*. *Predictive validity* är en typ av *criterion validity* och diskussionerna presenterade i 4.1.8 indikerar även användbarheten av *predictive validity* specifikt. Diskussion förs kring *false negatives* i 4.1.6 där respondent 15 pratar om att man potentiellt går miste om bra personal. *False positives* berörs av respondent 11 i 4.1.7 som berättar om de som utbildats till pilot men sedan inte klarar av det fysiska arbetet. Dessa diskussioner indikerar att även den tredje teoretiska konstruktionen är användbar vid utvärderingen av Isokai-testet.

Vi anser att det finns indikationer på att de olika teoretiska delarna i Payne & Harveys ramverk som vi presenterar i vår teoretiska referensram är användbara för att utvärdera selekteringstest baserade på fysiska attribut. Detta eftersom militär personal från Försvarsmakten (varav två är specialister på rekrytering) ansåg att faktorerna var relevanta att diskutera i relation till Isokai-testet

Slutsatsen ovan i denna analys kan dock ifrågasättas. Det kan anses problematiskt att vi bedömer användbarheten av enskilda faktorer i ett teoretiskt ramverk på det sättet som görs i denna studie. Faktorernas användbarhet för att utvärdera ett redan existerande selekteringstest bedöms med hjälp av kvalitativa intervjuer utifrån huruvida intervjupersonerna för diskussioner relaterade till faktorerna eller ej. Problemet är att vi som intervjuare har makt när det kommer till att styra diskussionen och därav bestämma vilka ämnen som är av intresse att diskutera. De vi intervjuar kan, i försök att vara tillmötesgående, välja att föra en lång diskussion om ett ämne de inte alls tycker är relevant för utvärderingen av Isokai-testet, enbart för att vi ställer en fråga relaterat till ämnet.

Risken för sådant beteende kan i detta fall påverkas av att vår semistrukturerade intervjumall i grunden är mer åt det strukturerade hållet. I våra intervjuer försöker vi ställa många olika frågor för att alla faktorer i ramverket ska ha möjligheten att diskuteras, eller avfärdas, åtminstone en gång. En intervjuperson som strävar mer efter att vara artig snarare än sanningsenlig kommer diskutera alla frågor (och således alla teoretiska delar) engagerat, även om de inte anser att de är relevanta.

Det kan också te sig problematiskt hur bedömningen av användbarhet går till då vi endast binärt fastslår om faktorerna är användbara eller inte. Att väga de olika faktorerna mot varandra för att kunna redogöra för deras inbördes och relativa användbarhet för utformningen eller utvärderingen av ett fysiskt selekteringstest hade varit intressant men är utanför denna studies omfång och inte heller möjligt baserat på den kvalitativa data vi samlat in. Med hänsyn till dessa begränsningar bör slutsatserna i detta avsnitt beaktas med kritiskt öga. Alla faktorer verkar vara användbara men ingen faktor kan bedömas vara mer användbar än någon annan.

## **5.2 Fenomen uppmärksammade vid appliceringen av Payne och Harveys ramverk på Isokai-testet som inte är en del av ramverket**

Under några av intervjuerna uppkommer diskussioner kring fenomen som Payne och Harveys ramverk verkar förbise. Dessa diskussionsområden berörs på en nivå som medför att en analys av dessa är av intresse. Trots att större delen av de ämnen som diskuteras under intervjuerna kan härledas till ramverkets faktorer är det vissa diskussioner som inte kan kopplas på ett tydligt sätt till ramverket. Denna begränsning utgör ett

tillkortakommande i ramverkets förmåga att på ett omfattande sätt bedöma fysiska selekteringstest, eller med andra ord, ett tillkortakommande i ramverkets användbarhet.

De fenomen som vi har svårighet att härleda till ramverket är den empiri som presenteras i 4.1.6 och 4.1.7. I båda fallen bedömer vi att ramverket faller kort när det kommer till att vägleda analysen av dessa två fenomen, två ämnen som vi anser vara viktiga vid utvärderingen av ett fysiskt selekteringstest.

### **5.2.1 Nuvarande eller potential styrka**

Flera av de intervjuade fokuserar på att Isokai-testet mäter en persons nuvarande prestation. De poängterar att detta utgör ett problem vid test av fysiska attribut eftersom det går snabbt att förbättra en individs fysiska förmåga senare under utbildningen. Vi anser att det kan vara omotiverat att Försvarmakten då diskvalificerar sökande, fullständigt eller för vissa befattningar, baserat på ett attribut som snabbt och enkelt förbättras. I synnerhet eftersom testet har visat sig utgöra ett hinder för kvinnor att etablera sig i Försvarmakten.

Fenomenet och dess till synes viktiga koppling till skapandet och utvärderingen av ett fysiskt selekteringstest uppmärksammas inte i Payne och Harveys ramverk. Detta trots att de fokuserar specifikt på just selekteringstest baserade på fysiska attribut. Det finns en aspekt och en faktor i ramverket som möjligtvis indirekt relaterar till diskussionen ovan. Aspekten ifråga är *performance standards* där, om man är medveten om fenomenet, möjligtvis kan kompensera för att man mäter kandidaternas styrkeprestation före träning och inte deras faktiska potential. Problemet består i att om man är omedveten om fenomenet och dess konsekvenser kommer detta vara svårt att uppmärksamma med hjälp av aspekten.

Man kan argumentera för att konsekvenserna av bristfällig hänsyn gällande nuvarande och potentiell prestation möjligtvis kan beröras i *criterion validity*. Om ett styrketest visar på låg *criterion validity* kan det bero på att kopplingen mellan nuvarande styrka och senare arbetsprestation är låg. Att isolera detta som orsak till låg *criterion validity* ter sig dock svårt för oss. En låg *criterion validity* kan lika gärna antas bero på en svag relation mellan styrka och senare arbetsprestation. Om detta fenomen ska uppmärksammas vid skapandet eller vid utvärderingen av ett fysiskt selekteringstest bör det nämnas explicit i ramverket. Med tanke på den fokus som ägnas av våra intervjupersoner åt detta fenomen anser vi att det är rekommenderat.

### **5.2.2 Ett selekteringstests kostnad**

Monetär kostnad är ett diskussionsämne i flera av studiens intervjuer (se avsnitt 4.1.4) I synnerhet förs en intressant diskussion kring att det kan vara relevant att inte stirra sig blind på testets direkta kostnader, något som presenteras i 4.1.7. Istället måste man även ta

hänsyn till de potentiella indirekta kostnader som kan uppstå av att använda ett till synes billigt men bristfälligt fysiskt selekteringstest.

Trots det uppmärksammas ett selekteringstests kostnad i begränsad omfattning i artikelns ramverk. Som kan utläsas i figur 1 tar artikeln upp faktorerna *resources, equipment samt administration* under aspekten *practical considerations*. Alla dessa faktorer kan tros relatera till diskussionen om ett tests kostnader, men innebörden av faktorerna, eller om de relaterar till kostnader, förklaras aldrig. I artikelns löpande text omnämns *practical considerations* i totalt två meningar:

*“Notwithstanding all of the theory, a PET must be able to be implemented feasibly and safely. This involves consideration of resources, equipment, time and personnel required, development of a clear and detailed set of test protocols and instructions regarding procedures for both test administrators and participants and associated safety considerations.”* (Payne & Harvey 2010)

Det enda stället i artikeln där kostnad omnämns är i en av ramverkets tabeller där skillnaderna mellan olika testkategorier summeras (återskapad i denna studie som tabell 2). I tabellen uppmärksammas de potentiella skillnaderna i kostnad mellan testkategorierna när de ska appliceras på flera jobbspecialiseringar. Kostnaden av ett test vid generell applicering är dock inget som tas hänsyn till i artikelns ramverk, än mindre direkta och indirekta kostnader.

Det står därför klart att ramverkets användbarhet för att utvärdera existerande fysiska selekteringstest begränsas av att ramverket till större del inte tar hänsyn till kostnaden av dessa. Ett fenomen som verkar vara av vikt vid utvärderingen av ett selekteringstest och därför bör vara en del av Payne och Harveys ramverk.

## 6. Slutsatser

### 6.1 Besvarande av forskningsfråga 1

*Vilka delar i Payne och Harveys ramverk är användbara vid utvärderingen av Försvarmaktens fysiska selekteringstest, Isokai-testet?*

Metoden vi använder för att besvara den första forskningsfrågan skapar begränsningar i vilka slutsatser som kan dras i analysen. Med tanke på de semistrukturerade intervjuernas uppbyggnad är det svårt att säga om intervjupersonernas diskussioner kring de olika delarna i Payne och Harveys (2010) ramverk medför att de är användbara för att bedöma Isokai-testet. Vidare kan inga slutsatser dras angående den relativa användbarheten mellan delarna.

Detta innebär att slutsatsen till den första forskningsfrågan saknar viss tyngd och att en mer uttömmande slutsats hade varit önskvärd. Det som kan konstateras är att alla delar i Payne och Harveys ramverk som studien berör verkar vara relevanta för att utvärdera ett selekteringstest baserat på fysiska attribut. Detta eftersom ingen intervjuperson avfärdade någon av diskussionerna relaterade till de olika delarna som irrelevanta.

## 6.2 Besvarande av forskningsfråga 2

*Vilka fenomen uppmärksammade vid appliceringen av Payne och Harveys ramverk på Försvarsmaktens fysiska selekteringstest, Isokai-testet, är inte en del av ramverket?*

Genom den empiriska undersökningen och analysen framgår det att Payne och Harveys ramverk inte tar hänsyn till två fenomen som vi anser vara av vikt vid utvärderingen av fysiska selekteringstest. Att fenomenen är viktiga baseras på att de uppmärksammas av flera intervjupersoner, varav två har omfattande erfarenhet och expertis inom området.

Det första fenomenet är faktumet att fysiska selekteringstest mäter nuvarande styrka och inte potentiell styrka. Payne och Harvey uppmärksammar inte detta faktum och inte heller de eventuella konsekvenser som kan uppstå på grund av det. Exempel på en möjlig konsekvens av fenomenet är att man underkänner fler kandidater som egentligen hade klarat av den efterföljande utbildningen. I Försvarsmaktens fall är det en potentiellt allvarlig konsekvens eftersom många kvinnliga sökande har förhindrats inträde i organisationen på grund av att deras styrka vid testtillfället ansetts för låg. För andra organisationer kan denna konsekvens tänkas utgöra ett problem för det kan innebära att man måste attrahera och testa fler sökande för fylla sina positioner än man annars hade behövt.

Det andra fenomenet som inte uppmärksammas i Payne och Harveys ramverk är kostnaden av ett selekteringstest. För Isokai-testet och Försvarsmakten är kostnaden en relevant diskussion på grund av budgetproblem (Zachariasson, Liebermann & Holm 2017). För andra organisationer behöver inte kostnadshänsyn vara av samma vikt men det är inte svårt att föreställa sig att fler organisationer som arbetar med fysiska selekteringstest upplever problem med budgeten. Ett exempel är den svenska polisen (Wetterqvist 2017).

Att Payne och Harvey inte har inkluderat de två ovan nämnda fenomenen anser vi vara ett tillkortakommande med hänsyn till fenomenens vikt vid utvärderingen av fysiska selekteringstest. Vi rekommenderar således att båda fenomen bör tas i beaktning vid applicering av ramverket för att skapa eller utvärdera ett fysiskt selekteringstest.

### **6.3 Praktiska implikationer**

Förhoppningsvis är implikationerna av vår studie att organisationer i större utsträckning kommer använda ett systematiskt tillvägagångssätt grundat i selekteringsteori för att skapa och utvärdera selekteringstest. Med andra ord att de applicerar Payne och Harveys ramverk samtidigt som de tar hänsyn till de två fenomen vi uppmärksammat, om situationen så kräver. I synnerhet hoppas vi att Försvarsmakten använder detta tillvägagångssätt för att utvärdera sin användning av Isokai-testet med tanke på att testet förhindrar kvinnors inträde i organisationen.

### **6.4 Studiens begränsningar**

Det finns flera begränsningar associerade med denna studie, och dess slutsatser, som är viktiga att erkänna. Något som redan uppmärksammats i viss mån är begränsningen i den utsträckning studien kan besvara forskningsfråga 1 baserat på kvalitativa semistrukturerade intervjuer, åtminstone i den form vi genomförde intervjuerna. Någon bedömning av ramverkets faktorerers relativa användbarhet var inte möjligt och mer än en indikation på att alla faktorer är användbara kunde inte etableras.

En annan begränsning i studien, som tagits upp i metodavsnittet, är att enbart två av de intervjuade var specialister inom rekrytering eller selektering. De andra intervjupersonerna hade visserligen både erfarenhet av Isokai-testet och av efterföljande utbildning (samt oftast anställning), men ingen djupare kunskap om selekteringstest. Bedömningen av Payne och Harveys ramverk, baserad på kvalitativa intervjuer, skulle antagligen varit mer utförlig och kvalificerad om fler av de intervjuade hade praktisk eller teoretisk expertis angående selektering. Det hade även varit av nytta att intervjua personer som var ansvariga för utformningen av Isokai-testet eller Försvarsmaktens rekryteringsprocess.

En tredje begränsning i studien är att studien genom besvarandet av sina två forskningsfrågor applicerar, bedömer och kritiserar Payne och Harveys artikel, men föreslår endast i begränsad utsträckning hur dessa tillkortakommanden i ramverket ska tacklas. Vi rekommenderar endast att de två identifierade fenomenen bör tas i beaktning och inkluderas i Payne och Harveys ramverk och därmed är studiens induktiva process svag. Det vore konstruktivt att i detalj föreslå ett nytt ramverk kompletterat med de insikter denna studie bidrar med. Detta har vi dock bedömt ligger utanför denna studies omfång och ambition.

### **6.5 Förslag till vidare forskning**

Med grund i ovan nämnda begränsningar presenterar det sig många möjligheter att bygga vidare på denna studies arbete. Att vidare undersöka forskningsfråga 1 och på ett mer definitivt sätt kunna etablera användbarheten i alla delar som Payne och Harvey

introducerar i sitt ramverk vore en naturlig fortsättning på denna studie. Vi rekommenderar kvalitativa intervjuer med personer som innehar praktiskt eller teoretiskt expertis inom selektering. Dessa intervjuer kan också möjligtvis involvera direkta diskussioner kring artikeln där dessa experter får ge sin åsikt angående artikelns användbarhet.

En annan naturlig utveckling med denna studie som bas är att introducera en reviderad version av Payne och Harveys ramverk där fenomen som kostnaden hos fysiska selekteringstest samt skillnaden mellan nuvarande och potentiell styrka uppmärksammas.

Ambitionen att introducera ramverk skulle även kunna tas ett steg längre. En möjlig fortsättning på denna studie är att introducera ett dedikerat ramverk för att utvärdera existerande selekteringstest, ett ramverk som kan användas parallellt med Payne och Harveys ramverk för att skapa nya selekteringstest. Detta ramverk kan sedan appliceras på existerande test som inte bedömts systematiskt från ett teoretiskt perspektiv tidigare, utan att Payne och Harveys ramverk först behöver anpassas för situationen. Det nya ramverket kan även användas för att kontinuerligt genomföra en teoretisk utvärdering av ett test för att bekräfta testets lämplighet över tid. Ramverket kan således säkerhetsställa att förändringar i testets kontext inte har gjort testets utformning mindre lämpligt (exempel inkluderar förändringar i jobbets utformning eller budgetförändringar i organisationen).

## 7. Referenser

### 7.1 Digitala källor

Bergman, E., 2017. *Diskriminering*. [Webbsida] Tillgänglig på: <http://www.ne.se/uppslagsverk/encyklopedi/lång/diskriminering> [Hämtad 2017-05-14].

Engelbrektsson, K., 2011. *En jämförande studie av isokinetiskt kontra isotont marklyft och drag till hakan*. BSc. Linnéuniversitetet. Tillgänglig på: <http://www.diva-portal.se/smash/get/diva2:424632/FULLTEXT01.pdf> [Hämtad 2017-04-12].

Försvarmakten, 2017a. *GMU*. [Webbsida] Tillgänglig på: <http://jobb.forsvarsmakten.se/sv/vagen-in/grundutbildning/om-gmu/> [Hämtad 2017-04-12].

Försvarmakten, 2015. *Antagningsprovningen på TRM, GU - Grundutbildning*. [Youtube-video] Tillgänglig på <https://www.youtube.com/channel/UC9013JWZWntBD6Zy9mVJQpA> [Hämtad 2017-04-12]

Försvarmakten, 2017b. *Befattningsutbildning*. [Webbsida] Tillgänglig på: <http://jobb.forsvarsmakten.se/sv/vagen-in/grundutbildning/befattningsutbildning/> [Hämtad 2017-04-12].

Försvarmakten, 2017c. *Förberedande officerskurser - FOK*. [Webbsida] Tillgänglig på: <http://jobb.forsvarsmakten.se/sv/vagen-in/ovriga-utbildningar/befals-och-officersutbildning/forberedande-officerskurser/> [Hämtad 2017-04-12].

Försvarmakten, 2017d. *Grundutbildning*. [Webbsida] Tillgänglig på: <http://jobb.forsvarsmakten.se/sv/vagen-in/grundutbildning/> [Hämtad 2017-04-12].

Försvarmakten, 2017e. *Personalsiffror*. [Webbsida] Tillgänglig på: <http://www.forsvarsmakten.se/sv/information-och-fakta/forsvarsmakten-i-siffror/> [Hämtad 2017-05-17].

Försvarmakten, 2017f. *Försvarmaktens årsredovisning 2016 Bilagor 1-3*. [Elektronisk källa] Stockholm: Försvarmakten. Tillgänglig på: <http://www.forsvarsmakten.se/siteassets/4-om-myndigheten/dokumentfiler/arsredovisningar/arsredovisning-2016/hkv-2017-02-21-bilagor-1-3-ar16.pdf> [Hämtad 2015-05-12].

Glaser, B.G. & Strauss, A.L., 1967. *The discovery of grounded theory: Strategies for qualitative research*. [E-bok]. Chicago: Aldine transaction. Tillgänglig på:

[http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser\\_1967.pdf](http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf) [Hämtad 2017-05-13].

Langlet, E., 2012. Omdiskuterat styrketest slår ut kvinnor. *Officerstidningen* [Elektronisk källa] Nr 1, 2012. Tillgänglig på: [https://www.officersforbundet.se/globalassets/pdf/officerstidningen/2011-2012/ot1\\_12\\_web.pdf](https://www.officersforbundet.se/globalassets/pdf/officerstidningen/2011-2012/ot1_12_web.pdf) [Hämtad 2017-05-12].

Larsson, H., Tegern, M., Monnier, A., Skoglund, J., Helander, J., Persson, E., Malm, C., Broman, L. & Aasa, U., 2015. Content validity index and intra- and inter-rater reliability of a new muscle strength/endurance test battery for swedish soldiers. *Plos one*, 10(7), <https://doi.org/10.1371/journal.pone.0132185>.

Newton, P.E., 2013. *Does it matter what validity means?*. University of Oxford, Department of education. Tillgänglig på: <http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2013/06/2013-Meaning-of-validity-Oxford-v4-slides.pdf> [Hämtad 2017-4-12].

Rekryteringsmyndigheten, 2017a. *Antagningsprövningen*. [Webbsida] (Uppdaterad: 2017-02-14) Tillgänglig på: <http://www.rekryteringsmyndigheten.se/malgruppsnavigering/om-du-skall-testas/forsvarsmakten-militar-grundutbildning/antagningsprovningen/> [Hämtad 2017-04-12].

Rekryteringsmyndigheten, 2016a, *Försvarsmaktens militära grundutbildning*. [Webbsida] (Uppdaterad 2016-09-15). Tillgänglig på: <http://www.rekryteringsmyndigheten.se/malgruppsnavigering/om-du-skall-testas/forsvarsmakten-militar-grundutbildning/> [Hämtad 2017-04-12].

Rekryteringsmyndigheten, 2016b. *Försvarsmakten - Militär grundutbildning*. [Webbsida] (Uppdaterad: 2016-04-06) Tillgänglig på: <http://www.rekryteringsmyndigheten.se/forsvarsmaktengmu> [Hämtad 2017-04-19].

Rekryteringsmyndigheten, 2010. *Registrering av längd, vikt och muskelstyrka*. [Webbsida] (Uppdaterad: 2010-12-30) Tillgänglig på: <http://www.rekryteringsmyndigheten.se/aktuella-ansokningar/forsvarsmakten-militar-grundutbildning/tester-och-undersokningar/registrering-av-langd-vikt-och-muskelstyrka/> [Hämtad 2017-04-19].

Rekryteringsmyndigheten, 2007. *Resultat från mönstringen - statistik från 2006*. [Elektronisk källa] Karlstad: Rekryteringsmyndigheten. Tillgänglig på: <http://www.rekryteringsmyndigheten.se/contentassets/50cec14487104438ad9c297d084587fd/statistik2006.pdf> [Hämtad 2015-05-12].

Rekryteringsmyndigheten, 2008. *Resultat från mönstringen - statistik från 2007*. [Elektronisk källa] Karlstad: Rekryteringsmyndigheten. Tillgänglig på: <http://www.rekryteringsmyndigheten.se/contentassets/50cec14487104438ad9c297d084587fd/statistik-monstring-2007.pdf> [Hämtad 2015-05-12].

Rekryteringsmyndigheten, 2009. *Resultat från mönstringen - statistik från 2008*. [Elektronisk källa] Karlstad: Rekryteringsmyndigheten. Tillgänglig på: [http://www.rekryteringsmyndigheten.se/contentassets/50cec14487104438ad9c297d084587fd/statistik\\_monstring2008.pdf](http://www.rekryteringsmyndigheten.se/contentassets/50cec14487104438ad9c297d084587fd/statistik_monstring2008.pdf) [Hämtad 2015-05-12].

Rönnerberg, L., 2017a. *Försvarsmakten*. [Webbsida] Malmö: NE Nationalencyklopedin AB. Tillgänglig på <http://www.ne.se/uppslagsverk/encyklopedi/lång/forsvarsmakten> [Hämtad 2017-04-19].

Wetterqvist, A., 2017. "Den miljard vi tillförs räcker inte" - rikspolischefen begär mer pengar av regeringen. *Dagens juridik*, [Webbsida] 3 Februari. Tillgänglig på: <http://www.dagensjuridik.se/2017/02/den-miljard-vi-tillfors-racker-inte-rikspolischefen-begar-mer-pengar-av-regeringen> [Hämtad 2017-05-13].

Zachariasson, H., Liebermann, A. & Holm, K., 2017. Så mycket pengar vill försvaret ha. *Svt nyheter*, [Webbsida] 2 mars. Tillgänglig på: <https://www.svt.se/nyheter/inrikes/sa-mycket-pengar-vill-forsvaret-ha> [Hämtad 2017-05-13].

## 7.2 Tryckta källor

American psychological association (APA), American educational research association (AERA) & National council on measurements used in education (NCME), 1954. Technical recommendations for psychological tests and diagnostic techniques, *Psychological Bulletin*, 51(2), s. 1-38.

Austin, J.T. & Villanova, P., 1992. The criterion problem: 1917-1992. *Journal of applied psychology*, 77(6), s. 836-74.

Barrett, G.V. & Luecke, S.B., 2004. Legal and Practical Implications of Banding for Personnel Selection. I H. Aguinis (ed.), *Test-score banding in human resource selection*. Westport, Conn: Praeger.

Bemis, S.E., 1968. Occupational validity of the general aptitude test battery. *Journal of applied psychology*, 52(3), s. 240-244.

Borman, W.C., Motowidlo, S.J. & Hanser, L.M., 1983. A model of individual performance effectiveness: thoughts about expanding the criterion space. Uppsats presenterad som en del av ett symposium: *Integrated criterion measurement for large scale computerized selection and classification, 91st annual American psychological association conference*, (Augusti).

Boxall, P. & Purcell J., 2000. Strategic human resource management: Where have we come from and where should we be going?. *International journal of management reviews*, 2(2), s. 183-203.

Boxall, P., Purcell, J. & Wright, P.M., 2008. *The Oxford handbook of human resource management*. Oxford: Oxford university press.

Brief, A.P. & Motowidlo, S.J., 1986. Prosocial organizational behaviors. *Academy of management review*, 11(4), s. 710–725.

Bryman, A. & Bell, E., 2013. *Företagsekonomiska forskningsmetoder*. Upplaga 2. Övers. B., Nilsson. Stockholm: Liber.

Campbell, D.T. & Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), s. 81-105.

Campbell, J.P., Asser, M.B. & Oswald, F. L., 1996. The substantive nature of job performance variability. I K. R. Murphy (ed.), *Individual differences and behavior in organizations*. San Francisco: Jossey-Bass.

Campbell, J.P., 1990. Modeling the performance prediction problem in industrial and organizational psychology. I M. D. Dunnette och L. M. Hough (eds.), *Handbook of industrial and organizational psychology*, vol. i. Palo Alto, Kalifornien: Consulting psychologists press.

Carver, R.P., 1974. Two dimensions of tests: Psychometric and edumetric. *American psychologist*, 29(7), s. 512-518.

Christensen, A.N., van Duin, C., Ahlström, L. & Gustafsson, E., 2016. *En robust personalförsörjning av det militära försvaret*. (SOU 2016:63). Stockholm: Regeringskansliet.

Cureton, E.E., 1965. Reliability and validity: Basic assumptions and experimental designs. *Educational and psychological measurement*, 25(2), s. 327-346.

- Delery, J.E. & Doty, D.H., 1996. Modes of theorizing in strategic human resource management: tests of universalistic, contingency, and configurational performance predictions. *The academy of management journal*, 39(4), s. 802-835.
- Dessler, G., 2013. *Human Resource Management*. Upplaga 13. Harlow: Pearson Education Limited.
- Dyer, W.G. & Wilkins, A.L., 1991. Better stories, not better constructs, to generate a better theory: A rejoinder to Eisenhardt. *Academy of management review*, 16(3), s. 613-619.
- Ebel, R.L., 1983. The practical validation of tests of ability. I: NAEP (National assessment of educational progress), *Conference on large-scale assessment*. Boulder, Colorado, 7-10 Juni 1982. Oxford, England: Blackwell publishing.
- Foster, S.L. & Cone, J.D., 1995. Validity issues in clinical assessment. *Psychological assessment*, 7(3), s. 248-260.
- Freebody, P. & Wyatt-Smith, C., 2004. The assessment of literacy: working the zone between 'system' and 'site' validity. *Journal of educational enquiry*, 5(2), s. 30-49.
- Försvarsmakten - Högkvarteret, 2015. *Antagningsordning för militär grundutbildning*, (FM2015-13373:1). Stockholm: Försvarsmakten.
- Grutter v Bollinger et al. (2003) 539 US Supreme Court. Docket N. 02-241.
- Guion, R.M., 1998. *Assessment, measurement and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Guion, R.M., 2011. *Assessment, measurement and prediction for personnel decisions*. Upplaga 2. New York: Routledge.
- Huselid, M.A., 1995. The Impact of Human Resource Management Practices on Turnover, Productivity and Corporate Financial Performance. *Academy of Management Journal*, 38(3), s.635-672.
- Jackson, A.S. & Hodgdon, J.A., 2000. Physical test validation for job selection. I: S. Constable & B. Palmer, eds. *The process of physical standards development*. State of the art report.

Karelitz, T.M., Jarvin, L., & Sternberg, R.J., 2010. The meaning of wisdom and its development throughout life. I: R.M. Lerner & W.F. Overton, eds. 2010. *The handbook of lifespan development: Cognition, biology, and methods* (Vol. 1). Hoboken, New Jersey: John Wiley & Sons. s. 837-875.

Kozlowski, S.W.J., & Klein, K.J., 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. I: K.J. Klein, & S.W.J. Kozlowski, eds. 2000. *Multilevel theory, research and methods in organizations*. San Francisco: Jossey-Bass.

Larsson, H., Tegern, M., Broman, L. & Harms-Ringdahl, K., 2011. *Isokai-testet som urvalstest till utbildning i Försvarsmaktens nya personalförsörjningssystem*. Stockholm: Karolinska Institutet.

LeCompte, M.D. & Goetz, J.P., 1982. Problems of reliability and validity in ethnographic research. *Review of educational research*, 52(1), s. 31-60.

Lee, B., Collier, P.M. & Cullen, J., 2007. Reflections on the use of case studies in the accounting, management and organizational disciplines. *Qualitative research in organizations and management: An international journal*, 2(3), s. 169-178.

Loevinger, J., 1957. Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), s. 635-694.

Organ, D.W., 1997. Organizational citizenship behavior: It's construct clean-up time. *Human performance*, 10(2), s. 85-97.

Organ, D.W., 1988. *Organizational citizenship behavior: The good soldier syndrome*. Lexington, Mass: Lexington books.

Payne, W. & Harvey, J., 2010. A framework for the design and development of physical employment tests and standards. *Ergonomics*, 53(7), s. 858-871.

Rotundo, M. & Sackett, P.R., 2002. The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, 87(1), s. 66-80.

Ryan, A.M., Mc Farland, L., Baron, H. & Page, R., 1999. An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel psychology*, 52(2), s. 359-392.

Saunders, M., Lewis, P. & Thornhill, A., 2012. *Research methods for business students*. Upplaga 6. Harlow, Essex: Pearson education limited.

Schmitt, N. & Chan, D., 1998. *Personnel selection: a theoretical approach*. Thousand oaks, California: Sage publications.

Schmitt, N., 2012. *The Oxford handbook of assessment and selection*. Oxford: Oxford university press.

Sharf, J.C., & Jones, D.P., 1999. Employment risk management. I: J.F. Kehoe, ed. 2000. *Managing selection strategies in changing organizations*. San Francisco: Jossey-Bass.

Shaw, S. & Weir, C., 2007. Examining writing: Research and practice in assessing second language writing (Studies in language testing 26). *TESL-EJ*, 11(4), s. 1-3.

Strauss, A. & Corbin J.M., 1990., *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, Kalifornien: Sage.

Terpstra, D. & Rozell, E.J., 1993. The relationship of staffing practices to organizational level measures of performance. *Personnel psychology*, 46(1), s. 27-48.

Wright, P.M., Boswell, W.R., Dunford, B., Snell, S.A., Gardner, T.M., Moynihan, L.M., Park H., Gerhart, B., Delery, J. & Allen, M.R., (2005). The relationship between HR practices and firm performance: Examining causal order. *Personnel psychology*, 58(2), s. 409-46.

Wright, P.M., Boswell, W.R., Dunford, B., Snell, S.A., Gardner, T.M., Moynihan, L.M., Park H., Gerhart, B. & Delery, J., 2001. Measurement error in research on human resources and firm performance: Additional data and suggestions for future research. *Personnel psychology*, 54(4), s. 875-902.

Yin, R.K., 2007. *Fallstudier: design och genomförande*. Upplaga 1:2. Övers. B., Nilsson. Malmö: Liber.

## 8. Bilagor

### 8.1 Bilaga 1: Intervjumall

#### Ska nämnas för intervjupersonen

Anonymitet

Uttalanden kommer inte kunna härledas till dig som person

#### Information om personen

Position inom FM?

Nivå av utbildning?

Antal år i FM?

Jobbat i olika delar av FM?

Värnplikt eller frivillig? (GMU eller inte)?

Vilket år gjorde du antagningstesterna? (Mönstrade/ansökte)

Ålder?

#### Uppvärmningsfrågor (Narrativ)

Om du minns, berätta gärna om när du gjorde rekryteringstesterna.

Det fysiska testet.

Var du nervös inför testerna?

Tränade du mycket inför testerna?

Gick det så bra som du önskade?

Fick du den befattning som du helst ville ha?

#### Huvudfrågor

Fråga 1: Vad tycker du om styrketestet?

Fråga 2: Tycker du att styrketestet är relevant för vad du gör/gjorde i FM?

Fråga 3: Tyckte du att styrketestet verkade relevant/vettigt då du gjorde det?

Fråga 4: Tycker du att förmågan som styrketestet mäter är

-väsentlig,

-användbar men inte väsentlig

-inte nödvändig

för att prestera väl inom Försvarmakten?

Fråga 5: Tycker du att styrketestet är ett bra test utifrån testets förmåga att:

Fråga 5.1: Mäta prestation noggrant och exakt?

Fråga 5.2: Inte vara diskriminerande?

Fråga 5.3: Vara pålitligt? Det vill säga man får samma resultat om man gör testet igen, eller med en annan maskin/på en annan plats?

Fråga 5.4: Att vara resurseffektivt och praktiskt utifrån vad man får?

Fråga 6: Tycker du att styrketestet är ett relevant test för att mäta nivån av den styrka som man behöver inom Försvarsmakten?

Fråga 6.1: Av den styrka man behöver inom den grundläggande utbildningen?

Fråga 6.2: Av den styrka du behöver i din vidareutbildning efter GMUn?

Fråga 6.3: Av den styrka du behöver i din nuvarande position?

Fråga 6.4: Av den styrka du behöver i andra positioner?

Fråga 6.5: Av den styrka som behövs generellt inom FM?

Fråga 7: Tror du att en person som får höga poäng på styrketestet vid antagningstesterna (7-9 poäng) blir en bättre tillgång för Försvarsmakten än en person som får låga poäng (2-4 poäng)?

Fråga 7.1: Tror du att personen blir en bättre soldat (skyttesoldat)?

Fråga 7.2: Tror du att personen kan bidra mer till att försvara landet vid händelse av krig?

Fråga 8: Tror du att styrketestet (och dess gränsvärden) är utformat för att välja ut de individer som Försvarsmakten faktiskt behöver eller att det är utformat för att välja ut individer som överensstämmer/påminner/är lika de individer som Försvarsmakten redan består av?

Fråga 9: Har du några idéer om hur man på ett annat sätt skulle kunna testa styrkan hos de personer som söker sig till Försvarsmakten?

Fråga 10: Tror du att det vore bra att ha ett test med flera moment som relaterar till olika vanliga aktiviteter inom Försvarsmakten? (Exempelvis: Klättra, marschera med vikt, dra en 80kgs docka, etc.)

Fråga 10.1: Vad tror du om ett Isokai-test med fler repetitioner istället för ett maxlyft som det är nu?

Fråga 11: Tror du att det kan vara en bra idé att redan vid detta tidiga stadiet ha olika styrketester beroende på vilken väg inom försvarsmakten man kan tänka sig att gå (vilka positioner man söker/är intresserad av)?

Fråga 12: Varför tror du att testet ser ut som det gör idag (varför testar FM personer med just ett ISOKAI-test och inte något annat)?

Fråga 12.1: Tror du att det beror på att testet har sett ut så länge?

Fråga 12.2: Att det finns motstånd mot att ändra testet?

Fråga 12.3: Att ingen tänkt på att utvärdera eller ändra testet?

Fråga 12.4: Att det inte finns något bättre test att genomföra?

Fråga 12.4.1. Att andra testmetoder är dyra/omständliga?

Fråga 13: Tycker du att man ska olika styrketester för kvinnor och män?

[Förklara om polisens arbete med sänkta krav för kvinnor på Isokai-testet]

Fråga 14: Tycker du att man skulle kunna göra något liknande i Försvarsmakten?

Fråga 15: Tycker du att man ska ha olika antagningskrav för kvinnor och män (exempelvis: för antagning till hemvärnet: 2/7 krävs för kvinnor, 3/7 krävs för män)?

## **8.2 Bilaga 2: Förteckning över figurer och tabeller**

Figur 1 - Aspekter och dess underliggande faktorer, introducerade av Payne och Harvey (2010)

Tabell 1 - Viktiga begrepp och dess definitioner

Tabell 2 - Styrkor och svagheter med respektive kategori av test

Tabell 3 - Information angående studiens intervjuer och intervjupersoner