

# As valid as it can be?

## Assessment of prior learning in higher education

**Tova Stenlund**



**Department of Applied Educational Science**  
**Educational Measurement**  
Umeå University  
No.6 2011

Department of Applied Educational Science  
Umeå University  
Doctoral thesis 2011

Cover art and design by Björn Sigurdsson

Printed by Print & Media, Umeå University  
Umeå, Sweden 2011

© Tova Stenlund

ISSN: 1652-9650  
ISBN: 978-91-7459-209-2

# Abstract

Assessment of prior learning (APL) represents the task to identify and acknowledge an individual's knowledge and skills regardless of how it has been obtained. In higher education this type of assessment is primarily used for the purpose of awarding access, credits or advanced standing. Because of the impact the results from APL have on the future working career for individuals claiming APL, it is of great importance that these result of APL is valid. The question of interest in this thesis is to what extent APL in higher education is a valid assessment.

The thesis is written in the field of educational measurement and comprises four papers and an extensive introduction with summaries of the papers. The most recent views of validity theory were used as the general theoretical framework in all papers, and all papers are concerned with APL in higher education. Study I reviews the research area of APL in higher education from a validity perspective. The general conclusion from the review is that the majority of the studies conducted in this area primarily provide theoretical rationales and theories for a variety of APL practices, and that there is a need for empirically based studies examining and evaluating validity of APL. Studies II, III and IV are empirical studies based on, and exemplified with, an APL scheme related to higher education in Sweden. Study II examines validity issues identified from claimants (individuals or students claiming APL) view of APL. The claimants' experiences from the specific APL scheme were examined using a questionnaire developed for that purpose. Conclusions drawn from the results are that possible threats to validity may exist in the administration of APL procedures, as well as in consequences of APL. Study III focuses on validity of admission decisions based on APL. The study examines decisions made by different higher education institutions for approximately 600 individuals applying for higher education based on their prior learning. The results show that the existing practice of APL needs improvements in order to obtain validity and trustworthiness in the decisions made in relation to APL. Finally, Study IV focuses on reliability in APL related to higher education. The study provides data of inter- and intra-rater reliability among judges in the specific APL scheme. The results show a lack of especially inter-rater reliability, and a conclusion is that reliability in this type of assessment should be further investigated.

The general conclusion from this thesis is that there is a need to take validity issues in APL seriously, and that APL in higher education may not be as valid as it could be.

*Keywords:* Access to Higher Education, recognition, accreditation, informal learning, validity, reliability, validation



# Acknowledgements

There are so many that have made it possible for me to reach this final goal, to complete my thesis. I sincerely hope that I do not fail to remember anyone (even if it seems to be in my nature) in this attempt to show everyone my gratefulness.

First, I would like to thank my supervisor Peter Nyström, for his feedback, support and encouragement. I would also like to thank him for his philosophy of giving me full autonomy (within certain boundaries of course) which has resulted in personal growth that is most valuable in this profession. I am also grateful to my assistant supervisor, Widar Henriksson, for his helpful guiding, and for his ambition to polish my rough assessment and measurement skills to a diamond finish.

For contributing with professional feedback and discussions of my papers, I would like to thank the National Graduate School of Educational Assessment (Nationella forskarskolan i pedagogisk bedömning).

Many thanks to the staff at CV in Skellefteå who have been involved in my studies for your support and great cooperation. I especially want to thank Gunnar Sundström, for giving me access to the data material from ValiWeb, and also for putting up with me and my critical mind these four years. Many thanks also to the assessors who took part in study IV, as well as to the participants in Study II who filled in the questionnaires, for giving some of their valuable time to help me conduct my studies.

I would also like to thank my colleagues and friends in the “unit” Educational Measurement at the Department of Applied Educational Science for an inspiring and an enjoyable environment to work in. Many thanks to my current and previous fellow doctoral students; Gunilla Näsström, Per-Erik Lyrén, Anna Sundström, Anna Lind-Pantzare for invaluable support; and my roommates in the doctoral-student-bunker, Peter Vestergren, Markus Strömbäck Hjärne and Daniel Arnesson, for interesting discussions (at least most of the time).

For reviewing my English, I would like to thank Gunnar Persson and Susanne Alger. Many thanks to Lotta Jarl for helping me with practical issues related to the thesis and dissertation, and Björn Sigurdsson for contributing with layout expertise.

Finally, my warmest thanks to all my friends and family, especially my husband for supporting me in everything I do, my children for never letting me lose track of what is important, my parents and my two sisters for teaching me how to be a somewhat decent grownup.

Tova Stenlund,  
Umeå, April 2011



# List of papers

This thesis is based on the following papers:

- I. Stenlund, T. (2010) Assessment of Prior Learning in Higher Education: a review from a validity perspective. *Assessment & Evaluation in Higher Education*. 35: 7, 783-797, First published on: 04 august 2009, ifirst.\*
- II. Stenlund, T. (in press) Threats to the valid use of Assessment of Prior Learning in Higher Education: Claimants` experiences of the assessment process. *Assessment in Education: Principles, Policy & Practice*.
- III. Stenlund, T. (2011) Validity of admission decisions based on Assessment of Prior Learning in Higher Education. *Accepted pending revisions, Assessment and Evaluation in Higher Education*.
- IV. Stenlund, T. (2011) Agreement in Assessment of Prior Learning related to Higher Education: an examination of Interrater and Intrarater Reliability. Manuscript submitted for publication.

All references to the papers in the thesis will be based on the enumeration used above.

---

\* Reprinted by permission of the publisher





## Swedish summary

Bedömning av reell kompetens (*assessment of prior learning* på engelska) är en företeelse som har uppmärksammats i stora delar av världen, däribland Sverige, under de senaste åren. Bedömning av reell kompetens handlar om hur människors kompetens, oavsett om det är formell (exempelvis utbildning) eller informell kompetens (exempelvis yrkeskompetens), kan identifieras, värderas och erkännas. Den exakta definitionen av bedömningen har diskuterats under åren och regeringens senaste förslag på definition lyder;

*"Validering är en process som innebär en strukturerad bedömning, värdering, dokumentation och erkännande av kunskaper och kompetens en person besitter oberoende hur de har förvärvats (DS 2003:23).*

Inom högre utbildning används den här typen av bedömning för både antagning och tillgodoräknande inom en kurs eller ett program. Eftersom resultatet av bedömningen får stora konsekvenser för individens framtida yrkesliv är det speciellt viktigt att bedömningen har god kvalitet. För att beskriva mätningars eller bedömningars kvalitet används konceptet validitet (enkelt förklarat är validitet att mäta det som är avsett att mätas) och innebörden i konceptet har förändrats under åren. Validitet som tidigare var uppdelad i olika typer av validitet är idag ett mer unisont koncept med begreppsvaliditet som det sammanhållande temat där fokus ligger på tolkningen som baseras på utfallet av mätningen eller bedömningen. Syftet med den här avhandlingen är främst att undersöka validiteten i bedömning av reell kompetens inom högre utbildning.

Avhandlingen innehåller fyra studier (I – IV) och en kappas. De fyra studierna i den här avhandlingen är alla fokuserade på validiteten, i synnerhet på hot mot validiteten, i den här typen av bedömning inom högre utbildning. En modern syn på validitet har genomgående använts som teoretisk grund för studierna i avhandlingen.

Studie I är en forskningsöversikt som undersöker forskningen som har bedrivits inom området mellan åren 1990-2007 och resultatet analyserades sedan utifrån ett modernt validitetsperspektiv. Det övergripande resultatet visade att majoriteten av forskningen inom det här området kan beskrivas som teoretiska studier, som exempelvis beskrivningar och jämförelser av procedurer för den här typen av bedömning, kritiska analyser och diskussioner om kvalitet i den här typen av bedömning. Resultatet visade också att det till stor del saknas empiriska studier som undersöker och utvärderar olika aspekter av validitet i bedömningen av reell kompetens.

Resultatet av studie I inspirerade till empiriska studier av en svensk procedur för bedömning av reell kompetens. Den här proceduren används bland annat för antagning till och tillgodoräknande i den svenska yrkeslärarutbildningen baserad på reell kompetens. Det finns ett antal olika yrkesområden att söka till och för att bli antagen till och få tillgodoräkna sig poäng i programmet så krävs det att den sökande har tillräcklig erfarenhet och kompetens inom yrket i fråga. I studie II, III och IV undersöks de olika delarna av den här specifika bedömningsproceduren och dess validitet.

Studie II undersöker proceduren ur deltagarnas (det vill säga de som har sökt till yrkeslärarutbildningen och grundat sin ansökan på reell kompetens) perspektiv. En enkät konstruerades för att få information om deltagarnas erfarenheter av bedömningsproceduren och cirka 330 av dem som sökt till yrkeslärarutbildningen genom den här specifika proceduren för bedömning av reell kompetens svarade på enkäten. Resultaten visade att det verkar finnas hot mot validiteten i såväl administrationen av bedömningen som i konsekvenserna av bedömningen. Exempelvis finns det skillnader mellan hur mycket hjälp de sökande har fått för att beskriva sin reella kompetens och en negativ konsekvens av bedömningen är att en del av deltagarna upplever att bedömningen saknar trovärdighet och rättvisa.

Studie III handlar om validiteten i de antagningsbeslut som tas grundande på bedömningen av reell kompetens i den här specifika proceduren. Studien undersöker cirka 600 individers antagningsbeslut till yrkeslärarutbildningen grundade på reell kompetens. Åtta olika universitet är representerade i studien. Resultatet visar bland annat att det finns skillnader mellan de olika universiteten som kan ifrågasätta validiteten och trovärdigheten i de här besluten och att förbättringar behövs för att besluten ska räknas som valida.

Studie IV fokuserar på reliabiliteten (det vill säga graden av överensstämmelse i en upprepad mätning av samma objekt) i bedömningen av reell kompetens inom högre utbildning. Studie IV undersöker både överensstämmelsen mellan bedömare (inter-rater reliability) och överensstämmelsen för bedömare som upprepar sin egen bedömning (intra-rater reliability). Resultaten visar att överensstämmelsen mellan bedömare är svag och sämre än överensstämmelsen för en upprepad bedömning av samma bedömare. En slutsats i den här studien är att reliabiliteten i den här typen av bedömning behöver undersökas ytterligare och att det krävs åtgärder för att den ska bli bättre.

Den generella slutsatsen i den här avhandlingen är att det verkar finnas brister i bedömningen av reell kompetens inom högre utbildning och att den inte är så valid som den borde vara. Det är viktigt att uppmärksamma betydelsen av kvalitet, i termer av validitet, i den här typen av bedömningar. Vidare studier är nödvändiga inom området och bedömningen har potential att nå en högre grad av validitet om förbättringar genomförs.

# Table of Contents

<b>INTRODUCTION.....</b>	<b>1</b>
RESEARCH SCOPE AND PURPOSE OF THIS THESIS .....	2
OUTLINE .....	2
<b>1. ASSESSMENT OF PRIOR LEARNING.....</b>	<b>3</b>
DEFINING APL.....	4
WHY ASSESS PRIOR LEARNING? .....	5
<i>The learning society and the concept of ‘lifelong learning’</i> .....	5
ASSESSING PRIOR LEARNING.....	7
<i>Theories and models of APL</i> .....	7
GUIDELINES OF APL.....	11
<b>1.2 WHAT IS VALIDITY AND HOW DO WE EVALUATE IT?.....</b>	<b>13</b>
EARLY PERSPECTIVES OF VALIDITY .....	13
TIMELY PERSPECTIVES OF VALIDITY .....	14
VALIDATION .....	17
<b>1.3 VALIDITY OF APL .....</b>	<b>21</b>
THE IMPORTANCE OF VALIDITY IN, AND VALIDITY STUDIES OF, APL.....	21
CHALLENGES RELATED TO VALIDITY OF APL.....	22
<b>2. MATERIALS AND METHODS.....</b>	<b>25</b>
<b>3. SUMMARY OF THE CONTRIBUTING STUDIES.....</b>	<b>29</b>
STUDY I .....	29
STUDY II.....	30
STUDY III.....	31
STUDY IV .....	33
<b>4. DISCUSSION .....</b>	<b>35</b>
MY CONTRIBUTION TO THE RESEARCH AREA OF APL .....	35
THE RESULTS RELATED TO GUIDELINES OF APL .....	37
LIMITATIONS AND GENERALISATIONS.....	38
FURTHER RESEARCH AND SUGGESTIONS FOR IMPROVMENTS .....	39
<i>As valid as it can be?</i> .....	40
<b>REFERENCES.....</b>	<b>43</b>



# Introduction

The changing economic and working climate has led to a growing awareness of the importance of lifelong learning. Today, few individuals will have a stable or single career pathway requiring education only at the career onset. The need to continually update transferable knowledge and skills as preparation for a number of working roles requires educational opportunities on a lifelong basis (Nyatanga, Foreman, & Fox, 1998). In the context of lifelong learning a great deal of attention is being given to assessment of prior learning. The assessment of prior learning (APL) represents the task of identifying an individual's knowledge and skills (regardless of how they were obtained) through suitable assessment, and of recording it as evidence and putting it to use (Evans, 1988). The practice of APL is built on the notion of saving resources. To avoid individuals being taught again what they already know is to save time and money for the individual as well as for society. Today, APL is used in a number of different contexts to provide individuals with opportunities for work or education they would not otherwise qualify for.

In the context of lifelong learning, policies to widen access to, and participation in, higher education have assumed increasing importance in Europe, as well as the rest of the world. Today, three main pathways are used to get access to higher education; Formal qualification, admission tests, and APL. In higher education APL is also used for the purpose of granting credits or advanced standing in a course or an education programme. Theory and practice of APL in higher education are based on the notion that experience promotes learning and that learning from experience can be made equivalent to academic learning (Shalem, & Steinberg, 2006). Even though, procedures of APL have been developed and are used in many countries, there is a lack of empirical research regarding the quality of this type of assessment procedure and its outcome. This is particularly important as APL may be regarded as high stake for the claimants. A test is high stake if the results have substantial consequences for the individual, and for APL the results definitely affect the career opportunities of the individual. In high stake assessments it is important to take quality issues particularly seriously and engage in extensive evaluation of the fully developed assessment in use (Moss, Girard, & Haniford, 2006).

Validity is a central concept in the evaluation of quality in assessments. Traditionally, validity refers to the extent to which an assessment actually assesses what it purports to assess, and was earlier seen as a property of a test or assessment and was also divided into different types of validity. Today, the most common view of validity is that it is a complex but unified concept and that it refers to the interpretation and use of a test or other

modes of assessment (Kane, 2006; Messick, 1989). Over the years several guidelines have been developed to ensure quality or validity in APL. However, to be able to ensure that an assessment is valid it is important that the theoretical evidence generated by following the guidelines is supported with empirical studies of validity (Shepard, 1993). When examining research related to APL it is obvious that there is a need for empirical validity studies investigating the degree of validity in this type of assessment.

## **Research scope and purpose of this thesis**

The present thesis is written in the field of educational measurement and assessment, and questions about what to assess, how to assess it, how to interpret the assessment, and how to secure the quality of these assessments, are particularly relevant in this research field. This thesis studies the quality of assessment of prior learning in higher education framed as validity, and one of the most important desires of any assessment is that it seeks to be valid.

The general aim of this thesis is to enhance the understanding of APL, and especially validity issues related to APL in higher education. Consequently, the thesis includes two different concepts that are particularly important, namely *assessment of prior learning* and *validity*. Both of these concepts will be thoroughly described, examined and related to each other in subsequent sections of this thesis.

## **Outline**

This thesis consists of two sections – an extensive introduction and a section including four papers. The first section includes four major parts. The first part aims at providing a definition and presentation of the two central concepts; assessment of prior learning and validity. Theories and models of both concepts are presented, and assessment of prior learning is related to validity. The second part in the first section contains a presentation of the method in the four studies included in this thesis. These studies are summarized in the third part. Finally, in the fourth part the main findings of the studies are discussed and suggestions for further studies are provided.

In the second major section of this thesis the four papers follow in numerical order. These four papers are targeted at scientific journals, and are freestanding pieces of research with separate discussions of relevant literature and methodological issues at hand.

# 1. Assessment of prior learning

Assessment of prior learning (APL) is widely used around the world and interest in APL and its procedures is increasing. In this part the concept of APL will be defined, some history and background of APL will be presented, and finally, theories and models of APL will be described.

The acronym APL refers to two terms that are relevant to identify before defining the entire idea of APL, namely *assessment* and *prior learning*. In the area of education *assessment* has been defined as the gathering, interpretation, and use of information to aid decision making, and in the context of APL *assessment* is a means of using different kinds of evidence to determine whether an individual claiming APL (claimant) has a required competence or not. The term *prior learning* has a central role in the concept of APL, and it refers to learning that has occurred before the candidate/student engages in the assessment process (Peruniak, & Welch, 2000). Prior learning usually includes at least three forms of learning; formal, informal and non-formal. Formal learning is sited in institutions dedicated to education or training, i.e. an organised and structured context (Bjørnåvold, 2000). Formal learning is structured via learning objectives or strategies facilitated by an instructor, teacher or trainer. Further, formal learning is intentional and leads to certification (Colley, Hodkinson, & Malcolm, 2006). Regarding informal and non-formal learning there seems to be no clear consensus on an exact distinction between the two (McGivney, 2006), and informal learning is sometimes considered to be a part of non-formal learning (Bjørnåvold, 2000). However, one definition is that informal learning takes place in everyday environments such as work, the home, the community and organisations (Colley, Hodkinson, & Malcolm, 2006). This type of learning is often referred to as experiential learning (Bjørnåvold, 2000), and is unstructured and often incidental. Non-formal learning takes place mainly in the workplace or community and voluntary settings, and is included in planned activities (Colley, Hodkinson, & Malcolm, 2006). The non-formal learning is both structured by a trainer, coach or mentor, and intentional, but it is normally not certificated. The concept of work-related or work-based learning is also used in the area of APL and refers to the learning taking place in a work context. This type of learning may be either informal or non-formal.

There also seems to be a lack of consensus on what term or label to use for the APL process around the world.

## Defining APL

APL has been given different labels in different countries. In UK, the accepted term is *accreditation of prior (experiential) learning (APEL)*, in the US and Canada it is *prior learning assessment (PLA)*, although Canada occasionally uses *accreditation of prior learning (APL)*, and in Australia, New Zealand and South Africa the term *recognition of prior learning (RPL)* is used. In France the terms *Validation des Acquis Professionnels* and *Validation des Acquis de l'Experience* are used, and a translation of the French word validation, *validering*, is usually used for APL in Sweden. In accordance with the French terms for APL the term *Validation of informal/non-formal learning* is frequently used in Europe, for example in CEDEFOP's (European Centre for the Development of Vocational Training) work in the area.

The definitions of these terms can vary from quite tight notions of APL as access, credit transfer or advanced standing in higher education, to views of APL as a reflective process with impact on the learning process (Austin, Galli, & Diamantouros, 2003; Donoghue, Pelletier, Adams, & Duffield, 2002; Gibbs & Angelides, 2004; Harris, 2006; Romaniuk & Snart, 2000). One example of a tight notion is the strict definition from the Australian Credit Transfer Agency's leaflet *Credit for Prior Learning* cited by Taylor (1996);

*...assessment of the learning which a student may have gained from his or her previous study and/or work experience, to establish whether this learning is equivalent to that which might have been gained in the university course in which he/she wishes to enrol (p. 282).*

Another example of a fairly strict definition of RPL is given by the South African Qualifications Authority, cited by Harris (1999);

*...the comparison of the previous learning and experience of the learner, howsoever obtained, against the learning outcomes required for a specified qualification and the acceptance for purposes of qualification that which meets the requirements (p. 136).*

In Gibbs and Angelides (2004) APEL is described with less rigour as a reflective process;

*APEL offers a mode of assessment which is dialectic and holistic. Through the reflection of practice learners review critical instances within their learning experience and from them disrupt their works at hand by revealing moments where these skills and the background of these actions create a discontinuity (p. 343).*

However, even if the terminology differs, it seems that there is consensus on the overall idea behind the different terms: to assess and acknowledge individuals' competence and knowledge regardless of how and where the competence was acquired. The development of APL procedures around the world makes it obvious that there is a need for this type of assessment and



that society values highly the acknowledgement of prior learning, in particular informal and non-formal learning.

### **Why assess prior learning?**

Assessment always goes beyond its obvious purpose of, for example, supplying a trustworthy judgement of the competence level of an individual, it is also a message about what is valued (Boud, 2000). Today we live in what is sometimes called a learning society and the concept of lifelong learning is strongly related to what we decide to assess.

### ***The learning society and the concept of ‘lifelong learning’***

The concept of lifelong learning is very much connected to APL and it is important to explain this relation further. The concept of lifelong learning, and the related concept of lifelong education, can be traced back to the early twentieth century (Tight, 1998). However, the concept did not become well-known until the 1960s. Lifelong learning is presented as a means to make it easier for individuals, organisations and nations to meet the challenges of an increasingly competitive world. Instead of education mainly restricted to childhood, lifelong learning or education was to last throughout life. Tight (1998) summarised the implications of lifelong learning. Firstly, as mentioned, it was to last the whole life for each individual. Secondly, it was to lead to systematic acquisition, renewal, completion and upgrading of competence to meet the constantly changing conditions of modern society, with the goal of promoting the self-fulfilment of each individual. Thirdly, to be successfully implemented it was to be dependent on individuals' increasing ability and motivation to engage in self-directed learning activities. Finally, it was to acknowledge all prior learning, including formal, non-formal and informal.

In accordance with the implications of lifelong learning, access to higher education or further education seems to be the main starting point for the idea of assessment of prior learning. According to Evans (2000) the phenomenon was first seen in the USA in the early 1970s. At first the main purpose of APL was social justice. At this time lifelong job security was breaking down and the possibility for adults to return to education was reachable with schemes like APL. In the USA a council for Adult and Experiential learning was established in 1974 and some of its main objectives were to develop and disseminate techniques for evaluation work and life experiences that can be given academic credits and to expand high-quality practice in assessment of adult learners' prior learning (Nyatanga, Foreman, & Fox, 1998). By 1974 more than a dozen institutions of post-secondary education were catering for 'non-traditional' students. The ideas included in

APL in the USA were exported to Britain in the late 1970s and early 1980s. In the mid- 1980s these ideas also found their way to the Canadian provinces of Ontario, British Columbia and Québec. The French government was concerned about the rising costs of providing higher education, and picked up the idea of APL from Quebec about the same time. By the late 1980s Australia, which embraced the idea from Britain, introduced its own version of APL. About the same time New Zealand, after extensive visits to both the UK and USA to find out about how APL procedures could be implemented, started to provide schemes for APL. In 1995 South Africa, with the election of the Mandela government, was also influenced by both the UK and USA and initiated APL schemes (Evans, 2000). In Europe the European Commission proclaimed 1996 as the year for lifelong learning, which resulted in an increasing interest in different activities related to APL in Sweden as well as most other European countries. Over time APL has exceeded its original university or higher education setting and has become an important part in the workforce development policy in many countries (Michelson, 1997). Michelson (1997) stated the aim of APL as fourfold;

*to provide individual workers with employment credentials; to enable employers to identify appropriately skilled workers; to help government and educational institutions identify needed areas for training and re-training; and to enhance the nation's economic edge at a time of global competition and technological change (p. 141).*

Even though interest in APL has increased in the last decades, and policies to widen access and participation have assumed increasing importance, there are still barriers in higher education related to APL. One of these barriers may be connected to the increasing concern about quality assurance and the ability of higher education institutions to maintain standards (Stowell, 2004). Stowell suggests that there could be “*an inevitable tension between a policy of widening participation and the maintenance of academic standards*” (p. 495). Another barrier is the universities' protection of their own status, i.e. universities' view of APL activity as low status (Michelson, 1997; Murphy, 2003; Taylor, 1996). A concern related to this barrier is raised by Gibbs and Morris (2001) when they argue for the safeguard of “*universities from becoming vessels for narrowly defined performance standards, unworthy of higher education.*” (p. 82). A third barrier is the cost. APL is considered to be time-consuming and expensive for a higher education institution (Taylor, 1996). Research also shows that many universities seem to lack real commitment to tackling the issue of widening access to higher education (Bateman & Knight, 2003; Colardyn & Bjørnåvold, 2004; Evans, 2000; Gallacher & Feutrie, 2003; Osborne, 2003; Pouget & Osborne, 2004), which obviously could have a connection to the barriers mentioned above. However, already in 1999 it was argued that the

barriers related to APL in higher education were weakening (Harris, 1999), and there seems to be an increasing interest in using and developing APL procedures in higher education.

## **Assessing prior learning**

APL in higher education is mainly used in two areas; firstly, as an evaluation of suitability and professional experience for entry to an education program which an applicant would not normally qualify for, and/or secondly, as an evaluation of a person's professional experience for the purpose of awarding credits or advanced standing, (Harris, 2006; Taylor, 1996). In general the aims of APL are the same as those of traditional assessment, i.e. to supply evidence primarily useful for selection, accreditation and qualification. However, formative (the assessment process in itself is a learning opportunity for the individual) and guidance (the APL should guide the individual in his or her choice of a future job or education) purposes are also frequently mentioned in the APL literature. In relation to the different purposes of APL different theories and models have been presented. These will be presented below.

### ***Theories and models of APL***

APL is generally seen as a holistic approach to the assessment of an individual's prior learning. As formal learning usually is well documented the main issue in this type of assessment is the task of assessing informal and non-formal learning. The accepted theorisation of APL is via experiential learning theory (Harris, 2006; Kolb, 1984; Peruniak & Welch, 2000), and APL is often based on the idea that learning from experience can be made equivalent to academic learning (Shalem & Steinberg, 2006). This view is also represented in Romaniuk & Snart's (2000) five assumptions underlying PLA (i.e. APL).

*First, that learning occurs across the lifespan. Second, learning takes place in various contexts, including formal, informal, and non-formal. Third, that formal learning is not necessarily of greater significance than learning gained through other contexts. Fourth, that formal learning objectives can be used to reliably assess learning gained through other contexts. And fifth, that when equivalent to formal learning, learning gained through other contexts should be recognized as so (p.31).*

A common approach in APL has been to engage the claimant (individual or student claiming APL) in some type of process in which they reflect upon their experiences in order to identify relevant prior learning (Whittaker, Whittaker, & Cleary, 2006). However, many different models of APL have

been developed during the last few decades and some of them have been identified in research related to APL.

Butterworth (1992) identified and described two models for APL, the *credit exchange model* and the *developmental model*. In the *credit exchange model* a claimant offers evidence of past achievements and credits or admission is awarded if this evidence indicates the necessary knowledge and abilities. Osman (2004) argues that this model is attractive to assessors and administrators because the process is seen as set of steps which are measurable and controllable. Assessment methods used in this model may include performance-based testing, examinations, and standardised tests (Osman, 2004). According to Osman (2004) this model of APL serves two purposes for the individual and the higher education institutions. First, for candidates who know what they want to study, and also know the value of their prior learning, this model offers a relatively straightforward way of gaining access to, or credits in, a specific education programme. Second, for those who are uncertain about the value of their prior learning and hold qualifications of uncertain value in higher education this model may serve a diagnostic purpose for both the institution and the individual claiming APL. In the *developmental model* the same kind of evidence is generated but there is also an additional element where the claimants are asked to evaluate and reflect on their prior experience and the learning derived from it. When awarding credits or admission, the assessor evaluates both the reflective account and the evidence. Trowler (1996) argues that the *credit exchange model* also involves some degree of reflection, and identifies a third model as the *credit exchange plus model*. It refers to the process of assessing or evaluating an applicant's claim of already having the competence needed to get access to or credits in a higher education programme. The competence can be shown either through demonstration or via a portfolio of evidence, which also involves the claimants' reflections on and identification of their prior learning. To contrast the *credit exchange model* and the *developmental model* Osman (2004) also suggested the *transformational model*. The *transformational model* recognises informal and non-formal learning on their own terms as valid academic knowledge.

Harris (1999) presents and analyses four models or perspectives of APL for the purpose of describing existing practices and to proposing possibilities for practices that would be able to make contributions to social inclusion; *procrustean*, *learning and development*, *radical*, and *Trojan-horse*. In the *procrustean model* only the aspects of an individual's prior learning which correlate with prescribed outcomes or standards are recognised. In this model, APL focuses on the demands from the labour market and consumer-oriented education, such as further education and vocational training. Instruments used in this model are performance testing, interviews, and evidence compiled into portfolios. The *learning and development model* is

used in higher education in what Harris calls newer institutions and areas such as social science and the professions “*where knowledge is less strongly classified and framed*” (p. 129). The method used for assessment in this model is mainly portfolio development. The claimants provide documentary evidence of relevant prior learning and reflective narratives in which they analyse their learning and also make comparisons with academic modes of learning. The articulated prior learning is then evaluated in terms of credits. The *radical model* is only found in contexts where the concern is radical social change. The prior learning is recognised through participation in social movements, with the view that the examination of experience from new perspectives leads to learning and knowledge. In this model prior learning is evaluated in terms of its emancipatory potential. The *Trojan-horse model* occurs in higher education where there is curriculum flexibility and weakening knowledge boundaries. The purpose is often to induce change by letting untraditional groups get access to higher education. In this model the informal and non-formal learning is evaluated in and of itself rather than solely in terms of its correlation with existing standards or curricula. Again portfolio seems to be the instrument used to organise and record prior learning.

Another effort to describe different models of APL was made by Andersson (2006). He suggests that APL could be divided into *convergent* and *divergent* assessments. The *convergent assessment model* is concerned with whether a person knows or can do something. This model is usually connected to criteria or standards controlling the assessment, i.e. evidence based. The *divergent assessment model* is concerned with what the person knows. In this model the prior learning is identified and evaluated through an exploring process to find out what the person knows know, i.e. process based. APL is also often connected with concepts such as summative and formative assessment, relating the convergent models to summative assessments and the divergent models to formative assessment (Andersson, 2006; Bjørnåvold, 2000).

It is argued that some of these models could be seen as two ends of a continuum rather than totally different models (Trowler, 1996), and most APL practice is also carried out as a mix of these models. However, to endorse simplicity I have made an attempt to picture the different models of APL presented above in relation to whether they could be considered to be process or evidence based (Table 1).

Table 1. Description of models of APL divided into process and evidence based.

	Process based	Evidence based
Purpose	Formative, Guidance	Selection, accreditation, qualification
Models	Developmental model Learning and developmental Trojan-horse Radical Divergent	Credit exchange model Credit exchange plus model Procrustean Convergent
Instruments	Portfolio development	Examination Performance or standardised tests Portfolio (dossier of evidence) Interviews

A common type of APL used in higher education is the evidence based model. In the process of APL in higher education the claimant is often asked to present a collection of evidence supported by narrative arguments, i.e. some type of portfolio, to demonstrate their prior learning (Joosten-Ten Brinke, Sluijmsmans, & Jochems, 2010; Klein-Collins & Hain, 2009; Lester, 2007). This collection is evaluated to identify whether the evidence can be accepted as credit, enable the student to receive advanced standing, or determine readiness for admission to a program or a specific course (Fjortoft & Zgarrick, 2001). Document analysis, authentic tests, knowledge tests, and interviews are other common instruments in this area (Joosten-Ten Brinke, Sluijmsmans, Brand-Gruwel, & Jochems, 2008; Lordly, 2007). Today, it is also common that this process, or parts of it, is computerized (Conrad, 2008; Sweygers, Soetewey, Meeus, Struyf, & Pieters, 2009). Even though the responsibility for making admission decisions and/or granting credits is placed on the higher education institution, an expert in the subject area of interest, for example a teacher or an external expert, usually judges the prior learning (CEDEFOP, 1997; Swegers, *et al.*, 2009).

As some type of portfolio seems to be the most commonly used instrument in this process it might be relevant to define this concept. There are, however, many different types of portfolios, but for the sake of simplicity I have chosen Trowlers' (1996) brief explanation of what defines a portfolio in this area. According to Trowlers a typical portfolio contains four elements. The first element is a summary of the learning claim. The second element of

the portfolio is a list of learning outcomes. The third element is a reflective writing describing the experiences and the learning resulting from the experience, and the fourth and final element is evidence to support the claim, for example employer certificates and testimonials.

Regardless of the type of model or instrument used in APL it is important that the assessment is of high quality, and several guidelines have been developed in order to help ensure that APL can be conducted with security and quality (See for example, Harris, 2000; Koenig & Wolfson, 1994; Nyatanga, Foreman, & Fox, 1998; Qualification Authority, 1993; Quality Assurance Agency for Higher Education, 2004; SIAST, 2000). The target audiences for these documents are often academic staff, policy-makers, planners and implementers of APL in higher education.

## **Guidelines of APL**

In Study I seven main principles for assuring high quality in APL in higher education were deduced from the different guidelines;

1. The procedure of APL should be clearly defined to claimants of APL, staff involved (such as assessors and examiners), and stakeholders. The criteria used to judge a claim for APL should also be included.
2. APL should only measure and evaluate what has been learned without consideration of the source of the learning.
3. The receiving academic institution is responsible for the assessment.
4. The APL process should be valid and reliable.
5. It should be treated with the same quality assurance procedures as other more traditional assessments.
6. The claimants for APL should receive guidance and support throughout the process, and the assessors involved should have the appropriate training.
7. Continuous quality improvement is necessary, e.g. the evaluation of policies and procedures.

In accordance with these guidelines the Council of the European Union (2004) has stated some common principles for APL practices in European countries. These have been divided into four main headings; Individual entitlements, Obligations of stakeholders, Confidence and trust and Credibility and legitimacy. Under the first heading it is argued that APL should be voluntary for the individual, with fair treatment and equal access for all. The APL procedure should also respect individuals' rights and privacy. Under the second heading it is stated that the stakeholders have the responsibility to establish systems for APL with the appropriate quality assurance mechanisms. The stakeholders should also provide information

and guidance about these systems. The third heading argues that APL procedures and criteria must be transparent, fair and underpinned by quality assurance. The fourth and final heading claims that APL procedures should be impartial and avoid any conflict of interest, and also that the professional competence of those who carry out the assessment should be assured.

The main issue in these guidelines and common principles is quality in the development and use of APL, and when it comes to quality the foremost important aspect is validity.



## **1.2 What is validity and how do we evaluate it?**

The uncomplicated definition of validity has traditionally been the extent to which an assessment or test measures what it claims to measure. In *Standards for educational and psychological testing* (AERA, APA, & NCME 2004) validity is defined as the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests or assessments, and is considered to be the most fundamental consideration when evaluating or developing an assessment. In the following text, after a short presentation of the early perspective of validity, the timely or modern perspectives and theories of validity and validation will be presented.

### **Early perspectives of validity**

Four types of validity were proposed, by the American Psychological Association in the mid 1950s; predictive, concurrent, content, and construct validity (Landy, 1986). Before this categorisation was introduced validity was simply considered to be the correlation between a predictor and a criterion. This view of validity was however not very helpful in developing basic understanding of what was being measured, and three validity types (predictive and concurrent validity are usually combined into criterion-related validity) content, construct and criterion-related validity were seen as a useful tool for validity research and discussions (Landy, 1986).

Content validity refers to the adequacy of sampling, i.e. how well the content of an instrument samples the content of the aspect to be assessed. Criterion-related validity refers to the correlation between assessment outcome and some kind of criterion (or external variable) for the assessment, and comprises the two subtypes' predictive and concurrent validity. Predictive validity indicates the extent to which the prior assessment performance can estimate an individual's future level on the criterion. Concurrent validity refers to the extent to which the outcome of the assessment estimates an individual's present standing on the criterion. Construct validity refers to the extent to which the content of an assessment or a measurement instrument are able to measure a theoretical concept or construct (Messick, 1989).

## **Timely perspectives of validity**

More recently the idea of a unitary concept of validity has been put forward by e.g. Messick (1989) and Kane (2006). In addition, the former view of validity as a property of the assessment has changed into a view according to which validity is evaluated in relation to the proposed interpretations and uses of the outcome of an assessment or test. Thus, the responsibility for the validity of the assessment is transferred from the test developer to the user (Gipps, 1995).

Messick argued that the focus in the unitary concept of validity should be on construct validity, in which content validity and criteria-related validity could be integrated. This is also visible in the six aspects of construct validity, that he claimed to be fundamental for all educational assessments; content, substantive, structural, generalisability, external, and consequential (Messick, 1989). The first aspect mainly relates to the traditional view of content validity. The second aspect is related to the specification of the construct's theoretical domain and its operational definitions. The third aspect is mainly related to methods for evaluating construct validity and also reliability. The fourth aspect is concerned with generalisability, i.e., the extent to which the assessment outcome and interpretations generalize across, for example, settings and groups. The fifth aspect includes the traditional view of how to examine criterion-related and construct validity. Finally, the consequential aspect refers to the appraisal of value implications of the assessment outcome as a basis for action as well as the consequences of the use of the assessment (these aspects of construct validity is more thoroughly examined by Eklöf, 2006). In Messick's validity chapter in the third edition of Educational Measurement (1989) he defines validity as

*an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (p. 13)*

Messick's framework is summarized in his four-fold matrix of validity. In this model he distinguishes two dimensions of validity. The first dimension represents the value of the assessment and is based on appraisal of either evidence or consequence of the assessment. The second dimension represents the function or outcome of the assessment. The function aspect is divided into interpretation and applied use of the assessment. Four aspects of validity are obtained when these dimensions are crossed (Table 2). Even though these four aspects seems to be clearly distinguished Messick argues that they are intertwined, and that it is important to take all four aspects into

consideration when discussing validity of, or validating, an assessment (Messick, 1989; 1995).

Table 2. Messick's four-fold matrix of validity (1989, p20)

		Function of the assessment	
		Interpretation	Use
Value of the assessment	Evidential basis	1. Construct validity	2. Construct validity + Relevancy/utility
	Consequential basis	3. Value implications	4. Social consequences

The first aspect is *the evidential basis of assessment interpretation* and represents evidence supporting the trustworthiness of the assessment outcomes. This aspect of validity points to evidence and rationales supporting the trustworthiness of the assessment outcomes, and has construct validity as its most important focus. According to Messick there are two main threats against construct validity; construct underrepresentation (the assessment leaves out things that should be included) and construct-irrelevant variance (the assessment includes something that should be left out). To avoid these threats a clear definition of the construct domain being assessed, content and cognitive specifications, is necessary to be able to judge how well the assessment matches this construct domain (Segers, Dochy, & Cascallar, 2003). It is also important to examine rival explanations for the observed performance. Furthermore, reliability, which has traditionally been viewed as a complement to validity, is in Messick's model viewed as part of the evidential basis of test interpretation (Gersten & Baker, 2002; Nyström, 2004). Reliability refers to the stability of the assessment and is usually explained as the degree to which the result or the outcome of an assessment can be replicated on a new occasion and by new assessors. Both instruments and assessors are factors that may cause reliability problems in assessments. Common reliability issues related to the instrument are poor instructions and/or unclear criteria for different levels in an instrument's scale (Kline, 2000). Difficulty in understanding how to use the instrument and/or the different scale levels may cause differences between occasions and/or assessors. Another source of assessment error is subjectivity in judging the quality of individual performance in assessment. In tests where objectively scored items (e.g. multiple choice items) can be

used, this threat to validity is not so serious, but in most types of judgements the outcome of the assessment rests on the assessors' ability to estimate the level of competence from a wide range of performances and evidence, which in turn allows for differences between assessors.

The second aspect of validity in Table 1 is *the evidential basis of assessment use*. This aspect is also construct validity, but its vital issues are relevancy, utility and accuracy of the assessment when it is used for different purposes. It is important to define the main purpose for which the results are planned to be used, as different purposes require different assessment designs. To allow users to draw adequately valid inferences from results the primary purpose must determine the most favourable design characteristics of the assessment (Newton, 2007). When the assessment has been designed with a primary purpose in mind, the operational problem will be how to ensure that results are not used for inappropriate purposes. However, since assessments often have more than one purpose it is important to state the multiple purposes of the assessment attention or we are in danger of damaging one or more of them (Boud, 2000).

The third aspect of validity in the four-folded matrix is *the consequential basis of assessment interpretation*. In this aspect the focus is on the appraisal of value implications of the meaning of assessment outcome. The assessment results, the way results are reported, feedback etc., have an influence on the conceptions of different participants such as the assessor and the assessed.

Finally, the fourth aspect of validity in Table 1 is *the consequential basis of assessment use* and this aspect focuses on appraisal of potential and actual social consequences. The question here is if the consequences harmonise with the intention of the assessment?

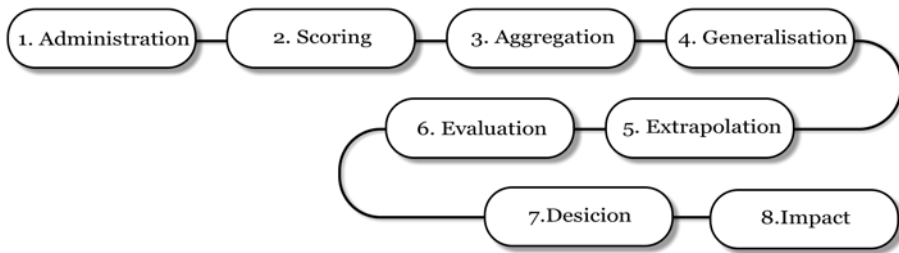
The consequential aspects in Messick's view of validity have caused discussions and not everyone agrees that these aspects are suitable to include in the validity concept (see for example Popham, 1997). However, the consequences of assessment were given attention already in the 1970s (Lyrén, 2009), and one could see it as if the discussion of consequences alone could cause a more thorough evaluation of assessments and tests, and also a more careful use of their outcome. Other problems with Messick's framework have also been articulated. The main complaints about his framework are that it does not help in the practical validation process, it is complex and difficult to understand, and further it is difficult to describe to lay audiences (Kane, 2006; Lissitz & Samuelsen, 2007; Shepard, 1993; Sireci, 2007). Nevertheless, his distinction of these four aspects is valuable when discussing the validity of an assessment and one could easily relate validity threats as well as theoretical and empirical evidence to his four-fold matrix. In Study I in this thesis Messick's matrix is used to describe the research area of APL, and another example of how to use the matrix is given

by Gersten & Baker (2002). However, as a consequence of the complaints described above, some models and theories of validation have been developed to aid the process of validation.

## **Validation**

Kane (1992) proposed an argument-based approach to validity which was thoroughly developed and described in the fourth edition of Educational Measurement (2006), in his chapter called *Validation*. In the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2004) validation is defined as a process that “involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations.” According to Kane the validity concept is connected to the argument proposed in the interpretation of the outcome of the assessment. The interpretation should be carefully examined and evaluated in the assessment’s arguments and in that way a reasonable validity is achieved in the assessment. The argument could be used to evaluate the assumptions underlying the interpretation and use of the outcome of the assessment. Four important inferences are described in Kane’s theory of validation; scoring, generalization, extrapolation and decision (see Lyrén, 2009 for a more thorough description and use of Kane’s theory). Further, Kane (2006) describes two different but closely related contexts where the term validation is found. The first context is when evidence is developed to support the interpretations and uses of a test or other modes of assessment, and the second context is when evidence is developed to support an assessment process. The main focus in this thesis is on examine threats to validity in the process of APL, or the degree of validity in the assessment of prior learning in higher education, rather than the development of evidence to support the assessment or the assessment process of APL. In the following part of the text a model examining threats to validity in assessments is presented.

Crooks, Kane and Cohen (1996) presented a model for the examination and evaluation of validity threats which was more elaborate than the four inferences found in Kanes later work (Kane, 2006). According to Crooks, Kane and Cohen (1996) validity is a multi-faceted concept and the practical validation of any assessment process requires a closer look at different links in the assessment process. Their model is a step-by-step approach aiming at evaluating the validity of educational assessments, and presents a chain of eight linked stages, including the four inferences or links described in Kane’s theory of validation. These eight stages describe different validity issues that have to be dealt with (Figure 1).



*Figure 1.* The eight links in the assessment procedure (Crooks, Kane & Cohen, 1996)

This model represents the whole assessment process from administration to impact of the assessment. In Study II in this thesis this model was used to examine threats to the valid use of APL from the claimants' points of view. There are several advantages of this model compared to the theory of Kane (2006). One advantage is that this model has the requested simplicity, i.e. it is very easy to understand, explain and use (for a practical example see Sundström, 2009). Another advantage is that this model is strongly related to the concept of assessment in the educational context, and it is also extended to involve all links in the assessment process. In the following text these links will be described, and examples will be given of possible threats to validity that are connected to the different links. Further, the links will be related to Messick's four-fold matrix.

The first link in this model is the administration of the assessment task. The outcome of the assessment can be significantly influenced by the procedures followed when administering and presenting the assignment. The performance can end up inappropriately low or high if proper procedures for administering and presenting the task are not followed.

The second, third and fourth link are scoring, aggregation and generalization of the performance in the assessment assignment. In other models of validation these are often merged into a single link referred to as reliability or generalisability. Possible threats against the valid use of assessment related to these links are for example, the lack of intra-rater or inter-rater reliability, i.e. the result depends on occasion or the assessor, or the threat of giving inappropriate weight to different aspects of the assignment.

The fifth link considers the extrapolation from the assessed field to a target field that holds all tasks relevant to the proposed interpretation. Major threats regarding this link are the well-known concepts of construct under-representation and construct irrelevant variance (Messick, 1989).

The sixth link is the evaluation of the performance or the forming of judgements. A possible threat against validity in this link is inappropriate judgements. The inappropriate judgement can involve both positive and negative bias. Positive bias occurs when the assessor judges the claimant more positively than the performance actually admits, maybe based on other information than the result of the assessment. Negative bias occurs when the assessor ignores unexpectedly high performance.

The seventh link is decision on actions to be taken in light of the judgements, and inappropriate standards or making poor pedagogical decisions are possible threats to validity related to this link. This link is closely related to the eighth link, which refers to the impact on the participants arising from the assessment process, interpretation, and decision. The decision to, for example, to give inappropriate feedback to the participants could have serious consequences for them. Threats such as positive consequences not achieved or serious negative impacts are related to the eighth link. Validity is also said to be reduced if the assessment is perceived to be unfair.

The threats in the first six links in the assessment model are strongly related to the evidential basis of validity, and mainly the first aspect of Messick's matrix, i.e. *the evidential basis of assessment interpretation*. Thus, the main threats are the aforementioned two threats to construct validity (construct-irrelevant variance and construct underrepresentation), but also threats related to reliability. The two final links in the model are strongly related to the consequential basis of validity. The threats representing the seventh and eighth links may be related to the third aspect, *the consequential basis of assessment interpretation*, as well as the fourth aspect of validity, *the consequential basis of assessment use*, of validity in Messick's matrix.

The literature related to APL is often prescriptive or descriptive and also seems to lack analysis and a focus on quality aspects of APL (Andersson & Fejes, 2005; Joosten-Ten Brinke, Sluijsmans, Brand-Gruwel, & Jochems, 2008). Assessing prior informal or non-formal learning, such as work experience, is a complex task, and since APL is used in higher education to gain admission and/or credits, concerns about quality in this assessment are particularly important. For APL, as well as other assessments, it is essential that central quality criteria, such as validity (including reliability), are met.





## 1.3 Validity of APL

Validity perspectives on APL in higher education suggest some challenges regarding current modes of assessment. These challenges will be investigated in this part of the thesis, but first there will be an explanation of why validity and validity studies of APL procedures are vital.

### **The importance of validity in, and validity studies of, APL**

*As with money, assessments can be understood as a code, providing information from holder to receiver. An individual applying for a job using assessments exemplifies this. Information as such is not enough, it must be presented in a specific code to be acceptable. As with money, assessments are valid in a predefined set of standard situations, e.g. in the labour market, within the hierarchy of an enterprise or in the system of education and training. Like money, assessment must also be based on some form of generalised value not only legal but also legitimate. The competences in question must be accepted as potentially valid/useful outside their narrow context of origin. (Bjornåvold, 2000, p.47)*

It is important that the APL procedure and its outcome are valid. If the outcome of the assessment is incorrect it will have consequences not only for the individual, but also for stakeholders and society. APL in relation to higher education is to be regarded as high stake for the individual, since the outcome of APL has far-reaching consequences for the individual's future working career. High quality is important for any assessment but the consequences of low quality are particularly severe when the assessment has far-reaching consequences for the individual participant. In high stake assessments it is vital that central quality criteria are met, and they also require an extensive evaluation of the fully developed assessment in use (AERA, APA, & NCME, 2004; Moss, Girard, & Haniford, 2006). Thus, to ensure validity in an assessment it is not enough to have theoretical rationales, such as for example following the guidelines for a valid APL assessment, it is also necessary to have empirical evidence to support validity (Shepard, 1993).

Validity studies are also critical for maintaining the credibility of any educational or psychological assessment (Sireci, 2003). Stakeholders need to be able to explain to the students and the rest of the society why, for example, the admission decisions should be relied upon (Shepard, 1993). To be able to ensure that the process of APL is valid as well as its outcome it is important to examine all aspects of validity. This is also important for its utility and exchange value in the labour market (Andersson, Fejes, & Ahn, 2004). As mentioned earlier in the text APL is viewed by higher education as a low-status activity, and one way to change this could be to conduct validity studies and to prove that validity issues related to the assessment are taken seriously.

Furthermore, it is important that validity studies examine the whole process of APL, i.e. assessment procedures, decisions made, and outcomes of APL. According to research, predictive validity seems to be high in APL in higher education (Cantwell, Archer, & Bourke, 2001; Donoghue, Pelletier, Adams, & Duffield, 2002; Marshall, & Jones, 2002; Rapley, Davidson, Nathan, & Dhaliwal, 2008), which could question the need for investigating the process of APL further. Still, predictive validity is only one aspect of the unified concept of validity, and therefore examining the predictive validity is not enough to ensure that the assessment process is valid, and further, the decisions made in APL could be unfair despite predictive validity.

### **Challenges related to validity of APL**

Even though APL has fundamental similarities to other assessments there are some challenges related to validity that are unique to the area of APL. Some of these validity issues are identified and discussed in Study I and will together with some additional challenges be described in this part of the thesis. Further, these challenges are described in relation to the two different approaches to APL, namely process and evidence based models of APL (see Table 1). (It is however important to remember that the models used often are a mix of these.)

In the process-based APL the most important challenge must be related to the aspects of validity referring to reliability. The character of the assessment makes it extremely difficult to evaluate reliability. It takes standardized methods to make comparisons between individuals and judges and to reach some degree of standardization one must have a set of relevant criteria (Andersson, 2006; Starr-Glass & Schwartzbaum, 2003). Since the main idea of the process-based APL is to value prior learning through an exploring process to find out what the claimant knows, this type of APL is not controlled by criteria or standards. However, Andersson (2006) argues that validity, disregarding the reliability aspect, may be high in a holistic assessment trying to explore what a claimant really knows. Another challenge in this approach of APL is the problem for claimants to be awarded academic credits or admission based on their own reflections on their learning (Harris, 1999). Harris pointed out that for this to happen it takes an institutional culture that is prepared to change. A problem regarding the claimants' reflections is that they are often unaware of, for example the depth and range of their own prior learning (Hager, 1998), and it is most likely that their own reflections on prior learning are affected by retrospective bias. Another challenge to this approach to APL is that it could easily be regarded as a boring process to "find oneself" and consequently not be treated seriously (Evans, 2000). The main purpose of this approach of

APL is usually formative and if, for example, the claimant for APL does not take the process seriously, or is treated in a way that decreases his or her motivation for learning, the argument related to the formative purpose may not be fulfilled.

There are also several challenges related to the evidence-based approach to APL. One challenge in this approach is how to relate informal or non-formal competence to criteria developed to measure formal learning. Hager (1998) analysed this issue and concluded that informal learning is typically different from formal education. Informal learning does not fit very well with the view of knowledge in formal education. Informal learning is also vastly contextual in contrast to formal education, which owns a privileged generality. Starr-Glass (2002) discussed validity in APL procedures in relation to the comparability of prior experiences and existing courses in higher education. He concluded that validity in APL needs to be revisited, and argued that APL should be viewed as a procedure that has predictive validity and allows us to see new connections between academic learning and the unique experience of the individual. Another issue in this approach is related to reliability. Even though, reliability in this type of APL could be high, as the assessment methods are more standardised and related to relevant criteria (Andersson, 2006), there are some reliability problems related to the use of portfolio. In contrast to tests that can use objectively scored items this type of assessment rests on the assessors' ability to estimate the level of competence from a collection of different evidence, which in turn allows for differences among assessors. A final validity challenge in APL is related to prior learning and experience. The claim that adults learn through experience is the most fundamental part of APL procedures. Even though this approach of APL is related to evidence of prior learning and specific criteria we have to ask what aspects of the experience need questioning and what parts of the experience have been misapprehended, ignored or omitted in recollection (Brookfield, 1998).

Referring to the research related to APL it seems as if there are some serious challenges related to the validity of APL, and the main focus of the studies of this thesis has been to examine threats to the validity of APL from both a theoretical perspective and an empirical perspective.



## **2. Materials and methods**

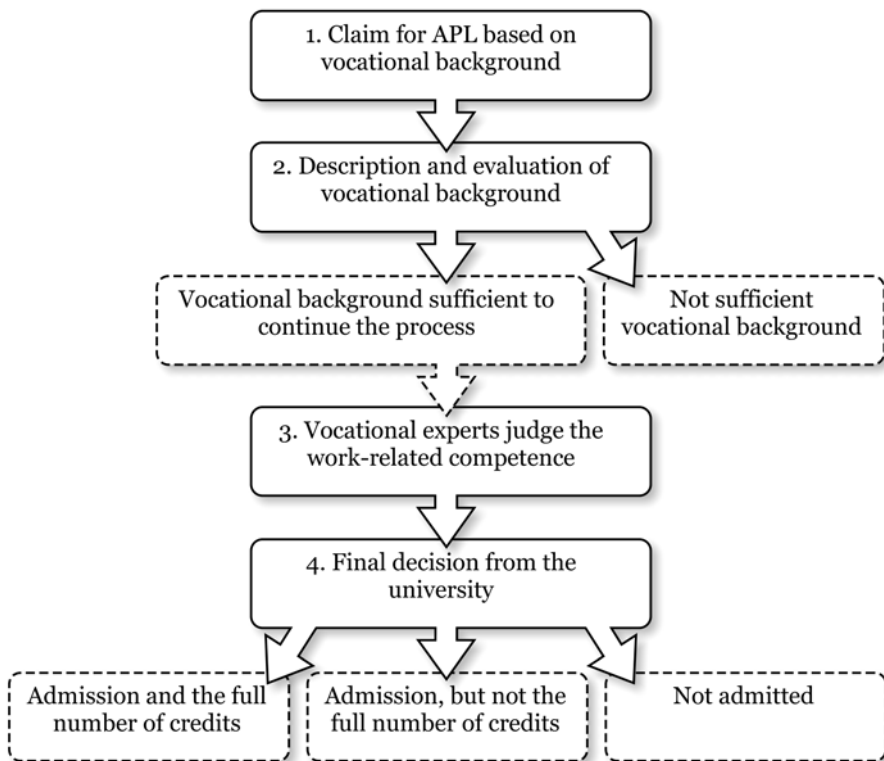
The main purpose of the studies in this thesis was to examine validity issues in relation to the APL process in higher education. All studies have been carried out with the modern concept of validity in mind.

Study I is a review of research in the area of APL from a validity perspective. The review was based on a search for relevant literature in several databases. As research in this area published before 1990 is relatively scarce, the search was limited to the years 1990 to 2007. The studies found were then analysed in relation to Messick's four-fold validity matrix. The review could be seen as the main starting point for my studies in this area and this thesis, and it was found to be most valuable for planning and carrying out the following studies. Study II, III and IV are empirically based studies examining different aspects of validity in relation to APL in higher education. These three studies are based on an APL-scheme used in relation to higher education in Sweden. The data used in these studies was collected in connection with this specific scheme. Before describing the materials and methods in these studies the APL scheme will be shortly described (a more extensive description of the APL scheme is found in Study III).

This specific APL scheme is developed and used for individuals applying for vocational teacher education programmes based on their prior learning. There are a number of different vocational areas to apply for and admission to such programmes requires that the applicant has sufficient professional experience in the vocational area in question. Based on the applicants' work-related competence they can be awarded admission and credits equivalent to up to one and a half years of study. This means that applicants can shorten their education to become certified vocational teachers by up to one and a half years. This APL scheme is a relatively representative example of APL in higher education. It is closely related to the model Trowler (1996) describes as the credit exchange plus model and also to the description of the most common procedure of APL in higher education described earlier in the section presenting theories and models of APL (p. 10).

The APL scheme is web-based and comprises four basic steps (see Figure 2). The first step is simply the applicant claiming to get his or her prior learning acknowledged in order to get access to and be awarded credits in the vocational teacher education programme. In the second step the applicant describes his or her work-related competence in a web-based instrument, and supports the narrative with some kind of evidence, if possible. The material is evaluated by staff working primarily with the web-based instrument used in this procedure. If the applicant seems to have the vocational background required the APL process proceeds. In the third step the applicant's collection of evidence supported by his or her narrative

argument is evaluated and judged by an external expert in the vocational field in question. The expert supplies judgements about the applicant's competence in nine different aspects of prior learning related to formal, informal and non-formal learning (the instrument used is more thoroughly described in Study IV). In addition to the nine separate judgements, the expert judgement also generates a recommendation to the receiving university concerning the decision of granting access and awarding credits to each applicant. In the fourth step the expert's judgements and recommendation, along with the applicant's collection of evidence, is evaluated by staff at the university that the applicant has applied to. The university makes the final decision to award the applicant credits and admission to the vocational teacher education programme or not.



*Figure 2.* Description of the APL scheme (this figure was originally presented in Study III)

The available data retrieved from this APL scheme includes the claimants' descriptions of their prior learning and also the outcome of each claim for APL, i.e. the expert judgements and the final decision. Eight different universities have used the web-based APL scheme and are represented in the data material. The sample includes claimants who applied for APL between May 2005 and August 2008. During this time a total number of 682 claims for APL were made, but the number of individuals participating was lower because some individuals claimed APL in more than one vocational area. A total of 632 individuals, 489 males (mean age, 41.87, S D = 7.93) and 143 females (mean age, 39.3, S D =7.64) made a claim for APL during this period. (32 males and 10 females claimed APL in two different occupations and 6 males and 2 females claimed APL in three different areas.) The total mean age was 41.2 (S D =7.93). The retrieved data also includes information about whether or not the participants have experience as vocational teachers.

In Study II, III and IV the different steps of this APL scheme were examined in relation to different validity issues, and the data retrieved from the scheme was used in Study II and III. In Study II the assessment procedure was examined from the participants' point of view. Thus, this study mainly focused on the first and second step of the scheme, but also on how they experienced the outcome of the APL. Besides data of the outcome of the APL procedure, data retrieved from a questionnaire was used to examine and analyse the claimants' experience and possible threats to validity related to the administration and consequences of APL. In Study III the focus was on the validity of the admission decisions, i.e. the fourth step in the APL scheme. In this study the recommendations from the vocational experts and the final decisions by the higher education institutions were used. Study IV was concerned with the reliability of the vocational expert judgements of the applicants' prior learning, thus the third step in the APL scheme described above. Two substudies were conducted in this study to be able to collect data of inter- and intra-rater reliability in this assessment.

### **Statistical methods**

The data materials in the studies have mainly been analysed with descriptive statistics, such as means, standard deviation and variability. Further, nonparametric techniques were used to analyse the results in Study II and III. As many statistical techniques assume that the distribution of scores on at least the dependent variable is normal, the outcome of the APL was tested for normality. The results suggested a violation of the assumption of normality, and to ensure some trustworthiness in the analysis of the results in Study II and III, they were conducted with nonparametric techniques. Correlation analyses (Spearman's rank order correlation) were made to find out if there is a significant relation between the final decisions by the higher education institutions and the experts' recommendations (Study III).

In Study II and III the data was explored to compare groups and to further examine if there are significant differences between groups with respect to the outcome of APL, using cross-tabulations and chi-square tests. In paper IV inter- and intrarater reliability was determined by using percentage of agreement and Cohen`s kappa statistic. The advantage with percentage of agreement is that it is simple to compute and explain, but it has some noteworthy weaknesses. The most important one is that it does not take chance into consideration, and therefore tends to inflate the degree of perceived assessor agreement, making it potentially misleading (Hayes & Hatch, 1999; Watkins & Pacheco, 2000). However, since Cohen`s kappa statistic, in contrast to percentage of agreement, takes chance agreement between assessors into account, it was used as a complement to percentage agreement. The kappa value ranges from -1 to 1, and values between .41 and .60 are suggested to be moderate, values above .60 substantial, and values below .40 poor.



### 3. Summary of the contributing studies

In this chapter the four studies in this thesis are summarised. The studies are related to one another in the sense that they all focus on validity of APL in higher education. Study I is a review of research in the area of APL in higher education examined from a validity perspective. Study II, III and IV are empirical studies that examine different validity issues in an APL-procedure used in higher education.

#### Study I

Study I in this thesis is a review of the research literature from a validity perspective. The review was conducted through a database search and was limited to the years 1990 to 2007 since the research before 1990 was rather scarce. The literature reviewed was then analysed and discussed with the help of Messick's theory of validity (Messick, 1995; 1989). The overall results revealed that the majority of research concerning APL presented in this period may be described as theoretical studies, such as descriptions and comparisons of APL procedures and perspectives, critical analyses of APL, and studies discussing and analysing quality issues in APL. However, the review pointed to a recent movement from theoretical studies to more empirically-based studies exploring the experience of claimants and assessors of APL and academic achievements of APL students.

The main conclusion from the validity analysis of the reviewed literature was that there is a need for further research in all four of Messick's validity aspects. In the *evidential basis of assessment interpretation* there is a lack of studies examining main threats against construct validity in APL, and the trustworthiness or reliability of judgements made in relation to APL is rarely considered. In the *evidential basis of assessment use* there is a need for empirical research investigating theoretical models of APL to examine how claimants experience or respond to different models. There is also a lack of studies examining different models to assess prior learning and in what way their results differ. In the *consequential basis of assessment interpretation* further research is needed to examine the degree to which APL influences the assessors' and claimants' perception of the value and relevance of the type of competence measured, and to what extent the assessors' and claimants' perceptions of the value and relevance of the assessment influence the outcome of the assessment. Regarding the *consequential basis of assessment use* it is important to examine positive as well as negative consequences and results of APL. It is crucial to examine if validity threats are responsible for the negative results and it is also important to examine if positive results of APL harmonise with the intention of the assessment.

## **Study II**

This paper focuses on the participants' experiences of APL. It is argued that participants in the assessment are an important source of information for the validation of the assessment. Study II was concerned with examining threats to validity in the links of administration and impact in Crooks, Kane and Cohen's (1996) model of assessment validation. A questionnaire, as well as data retrieved from the APL scheme (the outcome of the APL), was used to examine the claimants' experience and the threats to validity in these two links. The questionnaire was administered to 589 applicants in the collected data material and in total 328 questionnaires were completed representing a response rate of 56%. The respondents were considered to be a relatively representative sample, and thus results from the questionnaire may be considered to be relevant. The questionnaire provided data on individuals' perceptions of the procedure and outcome of the APL scheme. The results were analysed in relation to threats in the administration and impact link and revealed the following main results.

Administration refers in this case to the conditions under which the claimants present their prior learning which may have implications for the validity in assessment interpretation and use. The results showed that even though the majority perceived that the instructions and guiding of the APL procedure were fairly good, other results in the study implied that the understanding of the process might not be as satisfactory. If the claimants do not understand the instructions or the instrument of APL their performance may end up inappropriately low. The results also revealed that there seems to be a lack of clarity concerning what was required to gain access to or credits in the vocational teacher education programme, and about how the decisions were made. If the criteria for the assessment are not clear for the claimants of APL, it could result in difficulty in knowing what it takes to be successful in the assessment, and the claimants can end up with an inappropriately low performance. There is also some variation in how many times the claimants were in contact with the staff administrating the instrument of APL. This may imply that some claimants receive more help and instructions than others, thus resulting in an inappropriately high performance compared to those who did not contact the staff as many times. The results of this study also showed that time may be a problem for the claimants; this may be regarded as a validity issue if the time causes them to perform more poorly.

The second link examined in this study is the impact on the participants arising from the assessment process, interpretation, and decision (link eight in the model). The results related to this link showed that many of the participants who did not receive any credits or the full number of credits expressed great disbelief in the APL procedure and its fairness. Further, an often expected positive consequence of APL is better self-confidence and a more positive view of participants' work-related competence. Such consequences were also reported by claimants who got a positive result in the APL. It is, however, also obvious that a negative result, in this case not receiving credits at all or not the full number of credits, may have the opposite effect on the claimants' self-confidence and view of their work-related competence, i.e. a serious negative impact. The results showed that half of the claimants who were not admitted to the education programme did not consider going through an APL procedure again, which also could imply the negative consequence where claimants exclude themselves from further learning opportunities.

The main conclusion drawn from the results is that possible threats to validity may exist in the administration of APL procedures, as well as in consequences of APL.

### **Study III**

Study III focuses on validity of admission decisions based on this type of assessment in higher education. The study examines decisions made by eight different higher education institutions for approximately 600 applicants who used APL in order to receive admission to, and credits in the vocational teacher education program in Sweden. To examine validity issues in the admission decisions the vocational experts' recommendations and the final decisions by the higher education institutions, and the relation between these, were analysed using the outcome of each claim for APL. Group differences (gender and experience as vocational teachers) related to the outcome of the final decisions and the vocational experts' recommendations were also examined. The results were analysed and presented in relation to a validity discussion.

The results showed that there are significant differences between the vocational expert recommendations and the final decisions made by the higher education institutions. The final decisions seem to be more generous, in terms of admitting and giving credits, compared to the expert recommendations. Further, the results showed differences among the admitting universities. Some higher education institutions only use two of the three possible outcomes, disregarding the expert recommendations, and there also seems to be a difference among the institutions regarding the degree to which they follow the experts' recommendations. This result

suggests that some of the institutions make unfortunate decisions related to validity threats such as construct-irrelevant variance or construct underrepresentation. In this case the result could indicate either one. Thus, the higher education institutions omit the expert recommendations to some degree and/or apply some other criteria that may be irrelevant for the assessment.

The results also indicate a significant difference between males and females and between applicants with or without experience as vocational teachers regarding the outcome of APL. It seems to be significantly more difficult for females to receive the maximum number of credits compared to males, a difference that cannot be explained by age or years of occupational experience in a satisfactory way. The decisions about admission made by higher education institutions also seem to favour the claimants who worked as vocational teachers at the time of the APL procedure. Furthermore, the results indicate that it is harder for females with experience of teaching vocational courses to get a positive recommendation from the experts compared to the females without experience as vocational teachers. If the institutions judge the experience as vocational teachers as more important than the actual vocational competence, it could be argued that the assessment is threatened by construct irrelevant variance. Thus, the higher education institutions omit the expert recommendations to some degree and/or apply some other criteria that may be irrelevant for the assessment. On the other hand, it could also be argued that the expert judgements could be threatened by construct irrelevant variance. This type of assessment rests on the experts' ability to estimate the level of competence from a collection of different kinds of evidence, which in turn allows for differences among experts, i.e. subjectivity. Further, if we assume that the groups in this study are equal, the result indicating that the experts seem to judge females more negatively than males could be related to the validity threat of inappropriate judgement, and it could be argued that the experts' judgements are either positively biased towards males or negatively biased towards females.

The conclusion in this study was that there is a possibility that validity issues, such as construct underrepresentation and construct irrelevant variance, positive or negative bias, threaten the validity of the admission decisions in relation to APL.

## Study IV

This paper focuses on reliability in assessment of prior learning (APL) related to higher education. In this paper two studies investigating reliability in APL are presented. These studies provided data of inter- and intra-rater reliability among assessors in the APL scheme examined in the present thesis.

In the inter-rater reliability study two vocational areas were represented, and in each vocation two independent vocational experts judged the same claims of prior learning (i.e. the applicants' submitted evidence and written descriptions of their prior learning). In one vocation 13 applicants were assessed by the two experts, and in the other eight applicants. In the intra-rater reliability study two vocational experts in two different vocational areas participated. To investigate intra-rater reliability it is required that the vocational experts repeat the judging process of the same claims of prior learning. In this study two different vocations were represented and one assessor from each vocation repeated their assessment. In the first vocation claims from nine applicants were judged again, and in the other vocation 12 claims were judged again. The result includes the total inter- and intra-rater agreements of the vocational experts in the two vocations and the inter- and intra-rater agreements on the nine different aspects of APL. The vocational experts judge the claimants' prior learning in nine different aspects, and each area is judged on a four-level scale. As the vocational experts' judgements could be divided into two categories, not approved (representing levels one and two) and approved with or without restrictions (representing levels three and four), the second part of the results presents the inter-rater and intra-rater agreement representing only the two categories, *approved* and *not approved*.

The examination of agreement between raters, the inter-rater reliability, indicates a lack of reliability in this specific APL scheme. The agreement is poor using all four levels in the instrument and moderate or just above moderate using only the two categories approved or not. It is expected that formal aspects of prior learning are easier to assess reliably compared to informal aspects. But, the result shows surprisingly low agreement between experts on the level of the claimants' formal education or qualifications between experts. This difference in agreement could strongly indicate a difference in interpretations of how to use the instrument, i.e. what level a qualification or a specific formal education should be at, or different interpretations of the criteria for the assessment.

The results from the intra-rater reliability study reveal a higher agreement level than the inter-rater reliability study. The result also shows a difference in agreement between vocations (poor vs. moderate), suggesting a difference between vocational areas in explicitness of the requirements in the vocation. However, when examining the level of agreement using only the two categories, approved or not, the agreement is substantial and the difference between the different vocations is less noticeable. When examining the different aspects of APL in the intra-rater reliability study the result shows a pattern similar to the one identified in the inter-rater reliability study.

The overall conclusion is that there are strong reasons for concern regarding reliability in APL procedures related to higher education.

## **4. Discussion**

The main purpose of this thesis was to examine validity of APL in the settings of higher education, and hopefully by doing so the purpose of enhancing the understanding of APL will also be fulfilled. All papers in this thesis are concerned with validity issues of APL in higher education. In paper I a review of research in the area of APL was conducted and the results were related to Messick's theory of validity. In paper II, III and IV different aspects of validity were examined using a specific APL scheme as an example. In the first part of this text my contribution to this research area will be presented, and next the guidelines of APL will be related to my results. After that the limitations and validity of the results in this thesis will be discussed, and finally some future directions and concluding remarks will be made.

### **My contribution to the research area of APL**

My first contribution to this research area was the review of the research of APL from a validity perspective. The review pointed out areas in which the research is limited and in need of further development, such as validity studies in all four aspects of validity in Messick's four-fold matrix. Further, the review pointed out strong research areas, such as theoretical discussions and the development of theoretical rationales. In addition to the description of weaknesses and strengths regarding validity studies in this area of research, the use of Messick's four-fold matrix also contributed to a valuable and fairly general discussion about the importance of different aspects of validity and related validity threats.

My next contribution to this area is a relatively systematic investigation of validity threats related to the different stages in a specific APL procedure, i.e. studies II, III and IV. It is important to understand that all gathered information in an assessment procedure may contain minor or severe errors which could lead to incorrect judgements or decisions (Crooks, Kane, & Cohen, 1996). The question of validity is mainly about recognising that there are many different things that could contribute to different kinds of errors, and also about recognising that it is not always possible to limit them (Wedman, *et al.*, 2007). However, the critical part is to be aware of the limitations in the assessment, and more importantly, that if there are errors in the gathered information the conclusions drawn from the outcome of the assessment may be incorrect. Each of Study II, III and IV contributes information about validity threats in the APL procedure, which could be related to different aspects of validity.

In Study II the first link in the assessment process, i.e. the administration of APL was examined. The results revealed a number of problems to consider when administrating APL. For example, this investigation revealed the importance of a standardized administration of APL to avoid individuals being treated differently, which could cause inappropriately high or, more importantly, inappropriately low outcomes of APL. These results are particularly important for the understanding of how different aspects in the early stages of the APL procedure could affect the outcome of APL. The information gathered in this study is also important when developing and improving APL, and for staff, such as counsellors, working with this part of the process. The assessors' judgement of the claimants' prior learning was examined in Study IV. Since the research of APL reviewed in Study I revealed that the problem of reliability or the trustworthiness of the assessment is very rarely examined, this is an essential contribution to this research area. The study examined agreement among assessors, i.e. inter- and intra-rater reliability, in the APL-scheme. In this type of APL it is not expected to reach a high level of reliability, mainly due to the multifaceted evidence to judge. However, the result of this study made it clear that reliability of APL must be further examined and that improvements are necessary to achieve an acceptable level of trustworthiness. This study also described what seems to be most problematic from a reliability perspective and gave suggestions for provoking discussion on inter-rater reliability that might be useful for educating assessors of APL. Study III examined the final decision made by the higher education institutions. This study identified and discussed issues of validity in relation to how different universities make the final decisions, such as to what extent they follow the experts' recommendation. This study contributes to the discussion regarding the credibility of the decisions, and also to the discussion of the importance of a common model of APL for all universities to avoid the risk that individuals might have a disadvantage depending on what university they apply for.

The results above may be considered to be a valuable contribution to discussions of the evidential basis of assessment interpretation and use i.e. the first and second aspects of validity in Messick's matrix. It is, however, important to address validity issues relating to the consequential basis of assessment interpretation and use as well (Messick, 1995). The results of Study I revealed that consequences of APL are rarely considered in the research on APL, and therefore, one of the most important contributions to the research field of APL is the results described in Study II. The examination of the impact on the participants arising from the assessment process, i.e. the final link in the assessment process, enhances the understanding of how the procedure and the decisions made in the APL context may affect individuals and their view of themselves as well as their view of the assessment. Most serious is the finding of unintended negative



consequences, i.e. the withdrawal or exclusion from further learning opportunities depending on insufficient feedback or lack of transparency in the assessment. Considering APL's mission to endorse lifelong learning this result is particularly distressing, since it indicates the opposite.

To sum up, my contributions to the research area of APL are mainly the identification, discussion, and to some extent examination of serious threats to validity in this type of assessment procedure and its outcome. To further contribute to this area the results of my studies will be briefly discussed in relation to guidelines of APL.

## **The results related to guidelines of APL**

As mentioned earlier in this thesis the Council of the European Union (2004) has stated some common principles for APL practices in the European countries. These principles cover four key areas. In the following text some of the principles in these four areas will be presented and related to the results in this thesis.

### *Individual entitlements*

One of the principles stated in this area is that APL should ensure fair treatment and equal access for all. Fairness is often related to the concept of validity and referring to the results in this thesis it is clear that there are some threats to a fair treatment in these procedures. For example, the lack of standardized methods regarding the administration of the APL process could cause disadvantages for individuals depending on how much help and guidance they received (Study II). Furthermore, the indication of group differences in Study III, i.e. the possibility that it is more difficult for females to receive admission to higher education, could also shed doubt on the fairness.

### *Obligations of stakeholders*

In the second area one of the principles states that the stakeholders have the responsibility for establishing systems for APL with the appropriate quality assurance mechanisms. It is important that higher education institutions take APL seriously to be able to achieve a reasonable quality in the APL procedures. The research reviewed in Study I showed that many universities seem to lack a real commitment to tackling the issue of widening access to higher education, mainly because APL is considered to be time-consuming and expensive for a higher education institution. It was also shown that this lack of commitment and the barriers of cost in time and money could lead to deficiencies in the quality assurance of the assessment.

### *Confidence and trust*

In the third key area of principles it is stated that the procedures and criteria of APL must be transparent and fair. The importance of this principle is most obvious referring to the results of Study II. The result in Study II showed that the APL procedure investigated was not transparent to the claimants, and consequently the assessment was perceived as unfair by those that had a more negative result (i.e. did not receive the maximum number of credits or were denied admission). If the procedure and criteria for the assessment were transparent the result of the APL would not be perceived as unfair to the same extent. Thus, the claimants would have more knowledge of what caused the results.

### *Credibility and legitimacy*

In the final area one of the common principles states that the professional competence of those who carry out the assessment should be assured. To be able to make valid judgements and decisions the assessors of APL need to have proper education. However, the result in Study III and IV indicate that there may be a lack of professional competence and that the assessors (vocational experts) as well as the staff working with the admission decisions at the higher education institutions need further education, which is serious since this assessment is high stake for the claimants.

To avoid many of the threats to validity discussed above it is important that guidelines such as this are not only developed but actually also considered by the higher education institutions when developing and using APL.

## **Limitations and generalisations**

Each of the studies presented in this thesis have some limitations that need to be mentioned.

Firstly, the review in Study I was limited to research conducted after 1990. It is possible that research before this time could have added some valuable information about the area. However, research related specifically to APL before 1990 was difficult to find, and the examined studies nevertheless gave a valuable and somewhat updated picture of the research area. Secondly, the investigation of validity threats related to the administration and impact of APL in Study II focused on the applicants' post hoc perceptions. It could be argued that the results might risk some retrospective distortion, and that participants might have had some difficulties remembering the process. This could possibly have been avoided in the part examining the threats to validity in the administration of the assessment, by, for example, conducting a study that examined the claimants' perceptions during the time they went through the APL. However, based on interviews made in the process of

developing the questionnaire and the involvement of claimants who completed the questionnaire, the distance in time between the APL and this study is not a major problem. A majority reported that they could remember the APL procedure quite clear. Thirdly, a limitation in Study III is the small number of females in the data material. This could be one important alternative explanation for the differences in the results between males and females. However, since the vocational areas in this data material are mainly traditionally male the number of females is representative of the target population. Finally, a limitation related to Study IV is the small amount of data. The small sample size in this study may suggest that some caution is called for in the generalisation of the results. This study could have been helped by adding a qualitative approach to examine the judgement process of the assessors to shed some light on their preparation and how they use criteria. However, regardless of the limited data material, the study gives strong reasons for further examining the assessors' agreement on this type of assessment. It is far more serious to draw conclusions from a small sample when the results are positive, i.e. indicating that the assessment is valid, because there could be problems with the assessment that do not become visible due to the small sample. On the other hand, the detection of serious problems from a small sample is a strong indication that something really could be wrong, which the study of a larger sample could confirm.

The APL scheme examined in Study II, III and IV in this thesis is considered to be a relatively representative example of APL in the area of higher education. Consequently, the results from these studies are expected to be relatively generalisable to other contexts of APL in higher education using this type of APL model. However, one cannot draw any conclusions about other models of APL.

## **Further research and suggestions for improvements**

Some suggestions for further studies in all four aspects of Messick's matrix have already been made in the four contributing studies of this thesis. In the following section, some of these proposals will be emphasized and some additional suggestions will also be made.

One interesting question for future research, related to the result in Study III, is what aspects influence and/or explain the outcome of APL? In the outcome of APL both formal and informal learning are likely to play important roles. The assessment is however often predominantly focussing on prior informal or non-formal learning. An interesting question for future research is what weight the different aspects of prior learning are given in the assessment, and what other aspects influence the outcome? Moreover, as indicated by Study III, gender differences in the outcome of APL, mainly related to the vocational experts' judgements, are worthy of systematic

investigation. Drawing on the results from Study IV it is also important to examine the reliability of APL further, if possible with an extensive data material to be able to draw reliable and generalizable conclusions. As mentioned above a qualitative attempt to examine the assessors of APL could enrich the information about how this type of judgement is made.

In addition to research related to validity a suggestion will also be made related to one specific part of the procedure of APL. In Study III and IV the expert judgement and the final decision made by the higher education institution are examined. There is however one judgement in this procedure that is not included in this systematic examination of an APL procedure, namely the initial judgement made by the staff guiding the claimants through the APL process, i.e. the counsellor's part in this process. The counsellors have an important role in this procedure, because they are the first to meet the claimants and they are responsible for either including or excluding a claimant from continuing the process. Thus, it is vital to examine their role in this procedure further.

It is obvious that there is a need for further research examining the validity of APL, but the studies in this thesis have also shown that there is a great need for improvements in this area.

### ***As valid as it can be?***

The studies in the present thesis investigating assessment of prior learning in higher education from a validity perspective indicate a number of problems related to validity in this type of assessment. The overall conclusion in Study I is that there is a lack of empirical studies focusing on the validity of practices assessing prior learning. This conclusion inspired further studies aiming at investigating a specific Swedish procedure for assessment of prior learning for the purpose of granting access to, and credits in, higher studies. The results identify several serious threats to the validity of this assessment process. So, in answering the question posed in the title of this thesis, therefore, the answer can only be "No, not yet!"

Many times the excuse for not making an effort to improve or develop high quality assessment, and instead regard it *as valid as it can be* is related to costs, both in time and money. Even though APL is built on the notion of saving resources for society as well as for the individual, this way to *save resources* is not preferable. Low quality APL could be expensive. As the role of assessments is to provide decision makers with correct and relevant information to be able to make valid decisions the consequences of not being given correct information may be that the aim to have *the right individual at the right place* is not attained. In turn this could lead to individuals needing additional education, i.e. to higher costs for the individual as well as society,

or to individuals possessing competence that society actually needs being excluded, or excluding themselves from further education.

To increase the value and validity of this type of assessment it is important to make improvements of the assessment process and instrument used. Suggestions for improvements have already been made in the four contributing studies of this thesis, and some of these suggestions will be emphasized and summarized below. The improvements will be described in relation to the instrument used in the APL process, the competence to use it, the outcome of the assessment and its use, the competence to make the final judgement, and finally the consequences of the assessment.

Concerning the instruments used in APL it is important to have a transparent and clear definition of the competence being assessed, such as explicit criteria for the different aspects of prior learning. It is also important to have a clear definition for different levels in the instrument such as approved for admission or not. To reach a high level of consensus among the assessors (subject experts, such as teachers or external experts) and between occasions the information and training of the assessors using the instrument need to be improved. Additionally, the validity of the assessment could be improved by educating the vocational experts about the nature of their judgement process, and about threats underlying a valid judgement process (AERA, APA, & NCME, 2004). Regarding the outcome of the assessment and its use it is important that the assessment is only used with the primary purposes in mind. Even if the obvious purpose of APL in higher education is to award admission or credits, a more implicit purpose in these procedures is to be able to predict future job behaviour. It is important to consider all purposes of the assessment in order not to endanger or damage one of them (Boud, 2000). Regarding the competence at the higher education institutions it is vital to make a conscious decision to implement good APL procedures, and also to reach a common understanding among the institutions regarding the APL policy. The individuals making the final decision should also be educated about the nature of their judgement process and threats underlying a valid judgement process to enhance the validity of the outcome of APL. A final suggestion for improvement is related to the consequences of APL. To avoid negative consequences, such as exclusion from further learning opportunities, it is important to make appropriate pedagogical decisions, such as providing the claimants with relevant feedback about missing areas of competence. To gain positive consequences of APL it is vital to make the assessment procedure more transparent to the claimants of APL. If claimants know the criteria for the assessment, their performance will improve.

Finally, some changes have been made related to the APL scheme examined in this thesis that could lead to an improved procedure. From January 2011 there are new directions for admission to the vocational

teacher education programme. One of the consequences of these new directions for the APL procedure is that the outcome of the APL related to the vocational teacher programme will have only two outcomes in the future, either to receive admission with the full number of credits or not to be awarded admission. Another change is that new explicit criteria for each vocational area have been or will be developed. From a validity perspective the changes in the APL process related to the vocational teacher education seem to be to the better since the issue of transferring informal or non-formal learning into formal credits disappears, and clearly stated criteria may enhance the possibility for a more fair and trustworthy assessment. However, when explicit statements of required outcomes or prior learning are made openly available as the foundation for assessment decisions the effort to create consensus among staff at the university and the vocational experts must increase, as the fairness of the decisions made depend on this (Winter, 1994). This is most critical if APL in this context is to survive its exposure to public inspection, by stakeholders as well as by students.

# References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2004). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Andersson, P. (2006). Different faces and functions of RPL: an assessment perspective. In P. Andersson & J. Harris (Eds.), *Re-theorising the Recognition of prior learning*. Leicester: NIACE.
- Andersson, P., & Fejes, A. (2005). Recognition of prior learning as a technique for fabricating the adult learner: a genealogical analysis on Swedish adult education policy. *Journal of Education Policy*, 20(5), 595-613.
- Andersson, P., Fejes, A., & Ahn, S. (2004). Recognition of prior vocational learning in Sweden. *Studies in the Education of Adults*, 36(1), 57-71.
- Austin, Z., Galli, M., & Diamantouros, A. (2003). Development of prior learning assessment for pharmacists seeking licensure in Canada. *Pharmacy Education*, 3(2), 87-96.
- Bateman, A., & Knight, B. (2003). *Giving credit. A review of RPL and credit transfer in the vocational education and training sector, 1995-2001*. Leabrook: NCVET.
- Bjørnåvold, J. (2000). *Identification, assessment and recognition of non-formal learning in Europe. Making learning visible*. Thessaloniki: Cedefop.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Brookfield, S. (1998). Against naive romanticism: from celebration to the critical analysis of experience. *Studies in Continuing Education*, 20(2), 127-142.
- Butterworth, C. (1992). More than one bite of APEL - contrasting models of accrediting prior learning. *Journal of Further and Higher Education*, 16(3), 39-51.

- Cantwell, R., Archer, J., & Bourke, S. (2001). A comparison of the academic experiences and achievement of university students entering by traditional and non-traditional means. *Assessment & Evaluation in Higher Education*, 26(3), 221-234.
- CEDEFOP. (1997) Identification, validation and accreditation of prior and informal learning. United Kingdom Report. Thessaloniki: Report Research.
- Colardyn, D., & Bjornávd, J. (2004). Validation of formal, non-formal and informal learning: policy and practices in EU member states. *European Journal of Education*, 39(1), 69-89.
- Colley, H., Hodkinson, P., & Malcolm, J. (2006). European policies on 'non-formal' learning. A genealogical review. In R. Edwards, J. Gallacher & S. Whittaker (Eds.), *Learning outside the academy. International research perspectives on lifelong learning*. New York: Routledge.
- Conrad, D. (2008). Building knowledge through portfolio learning in prior learning assessment and recognition. *The Quarterly Review of Distance Education*, 9(2), 139-150.
- Council of the European Union. (2004). *Common European principles for the identification and validation of non-formal and informal learning*. European Commission. [www.europa.eu.int/comm/education/index\\_en.html](http://www.europa.eu.int/comm/education/index_en.html)
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy and Practice*, 3(3), 265-285.
- Donoghue, J., Pelletier, D., Adams, A., & Duffield, C. (2002). Recognition of prior learner as university entry criteria is successful in postgraduate nursing students. *Innovation in Education and Teaching International*, 39(1), 54-62.
- Eklöf, H. (2006). *Motivational beliefs in the TIMMS 2003 context* (Doctoral Thesis). Umeå: Umeå University.
- Evans, N. (1988). *Handbook for the assessment of experiential learning*: Learning From Experience Trust.



- Evans, N. (Ed.). (2000). *Experiential learning around the world. Employability and the global economy*. London: Jessica Kingsley Publisher.
- Fjortoft, N. F., & Zgarrick, D. P. (2001). Survey of prior learning assessment practices in pharmacy education. *American Journal of Pharmaceutical Education*, 65(Spring), 44-53.
- DS 2003:23. *Validering m.m. - fortsatt utveckling av vuxnas lärande*. Stockholm: Regeringskansliet.
- Gallacher, J., & Feutrie, M. (2003). Recognising and accrediting informal and non-formal learning in higher education: an analysis of the issues emerging from a study of France and Scotland. *European Journal of Education*, 38(1), 71-83.
- Gersten, R., & Baker, S. (2002). The relevance of Messick's four faces for understanding the validity of high-stakes assessment In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students. Validity, technical adequacy, and implementation*. New Jersey: Lawrence Erlbaum Associates.
- Gibbs, P., & Angelides, P. (2004). Accreditation of Knowledge as Being-in-the-world. *Journal of Education and Work*, 17(3), 333-346.
- Gibbs, P., & Morris, A. (2001). The accreditation of work experience: whose interest are served? *The Learning Organization*, 8(2), 82-88.
- Gipps, C. (1995). *Beyond testing. Towards a theory of educational assessment*. London & New York: The Falmer Press.
- Hager, P. (1998). Recognition of informal learning: challenges and issues. *Journal of Vocational Education and Training*, 50(4), 521-535.
- Harris, J. (1999). Ways of seeing the recognition of prior learning: what contribution can such practices make to social inclusion? *Studies in the Education of Adults*, 31(2), 124-142.
- Harris, J. (2000). *RPL: Power, pedagogy and possibility. Conceptual and implementation guides*. Pretoria: Human Sciences Research Council.
- Harris, J. (2006). Introduction and overview of chapters. In P. Andersson & J. Harris (Eds.), *Re-theorising the recognition of prior learning*. Leicester: NIACE.

- Hayes, J. R., & Hatch, J. A. (1999). Issues in Measuring Reliability: Correlation versus percentage of agreement. *Written Communication, 16*, 354-367.
- Joosten-ten Brinke, D., Sluijms, D. M. A., & Jochems, W. M. G. (2010). Assessors' approaches to portfolio assessment in assessment of prior learning. *Assessment & Evaluation in Higher Education, 35*(1), 55-70.
- Joosten-Ten Brinke, D., Sluijsmans, D. M. A., Brand-Gruwel, S., & Jochems, W. M. G. (2008). The quality of procedures to assess and credit prior learning: Implications for design. *Educational Research Review, 3*, 51-65
- Kane, M. T. (1992). An argument-based approach to validation. *Psychological Bulletin, 112*(3), 527-535.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (Fourth ed., pp. 17-64). Westport: Praeger Publishers.
- Klein-Collins, B., & Hain, P. (2009). Prior learning assessment: How institutions use portfolio assessments. *The Journal of Continuing Higher Education, 57*(3), 187-189.
- Kline, P. (2000). *Handbook of psychological testing* (2 ed.). London: Routledge.
- Koenig, C., & Wolfson, G. (1994). *Prior learning assessment in British Columbia. An orientation for postsecondary institutions*. Victoria: British Columbia Ministry of Skills, Training and Labour.
- Kolb, D. A. (1984). *Experiential learning. Experience as the source of learning and development*. Engelwood Cliffs, New Jersey: Prentice Hall P T R.
- Landy, F. J. (1986). Stamp collection versus science: Validation as hypothesis testing. *American Psychologist, 41*.
- Lester, S. (2007). Professional practice projects: APEL or development? *Journal of Workplace Learning, 19*(3), 188-202.
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*(8), 437-448.

- Lordly, D. (2007). Dietetic prior learning assessment: Student and faculty experiences. *Canadian Journal of Dietetic Practice and Research* 68(4), 207-212
- Lyrén, P.-E. (2009). *A perfect score. Validity Arguments for college admission tests*. Umeå University, Umeå.
- Marshall, G., & Jones, N. (2002). Does widening participation reduce standards of achievement in postgraduate radiography education? *Radiography*, 8, 133-137.
- McGivney, V. (2006). Informal learning. The challenge for research. In R. Edwards, J. Gallacher & S. Whittaker (Eds.), *Learning outside the Academy* (pp. 11-23). New York: Routledge.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Michelson, E. (1997). The politics of memory: The recognition of experiential learning. In S. Walters (Ed.), *Globalisation, Adult Education & Training. Impacts & Issues*. London: Zed Books Ltd.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30, 109-162.
- Murphy, A. (2003). Is the university sector in Ireland ready to publicly assess and accredit personal learning from outside the academy. *European Journal of Education*, 38(4), 401-411.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2), 149-170.
- Nyatanga, L., Foreman, D., & Fox, J. (1998). *Good practice in the accreditation of prior learning*. London: Cassell.
- Nyström, P. (2004). Reliability of educational assessments: the case of classification accuracy. *Scandinavian Journal of Educational Research*, 48(4), 427-440.

- Osborne, M. (2003). Policy and practice in widening participation: a six country comparative study of access as flexibility. *International Journal of Lifelong Education*, 22(1), 43-58.
- Osman, R. (2004). Access, equity and justice: Three perspectives on recognition of prior learning (RPL) in higher education. *Perspectives in Education*, 22(4), 139-145.
- Peruniak, G. S., & Welch, D. (2000). The twinning of potential: Toward an integration of prior learning assessment with career development. *Canadian Journal of Counselling*, 34(3), 232-245.
- Popham, J. (1997). Consequential validity: Right concern - wrong concept. *Educational Measurement: Issues and Practice*, Summer 1997, 9-13.
- Pouget, M., & Osborne, M. (2004). Accreditation or validation of prior experiential learning: knowledge and savoirs in France - a different perspective? *Studies in Continuing Education*, 26(1), 45-65.
- Qualification Authority. (1993). *The recognition of prior learning. Quality assurance in education and training* (Report). Wellington: New Zealand Qualification Authority.
- Quality Assurance Agency for Higher Education. (2004). *Guidelines on the accreditation of prior learning*. Gloucester: The Quality Assurance Agency for Higher Education.
- Rapley, P., Davidson, L., Nathan, P., & Dhaliwal, S. S. (2008). Enrolled nurse to registered nurse: Is there a link between initial educational preparation and course completion? *Nurse Education Today*, 28, 115-119.
- Romaniuk, K., & Snart, F. (2000). Enriching employability: the role of prior learning assessment and portfolios. *Journal of Workplace Learning: Employee Counselling Today*, 12(1), 29-34.
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimizing new modes of assessment: In search of qualities and standards*. Dordrecht: Kluwer Academic Publishers.
- Shalem, Y., & Steinberg, C. (2006). Portfolio-based assessment of prior learning: a cat and mouse chase after invisible criteria. In P. Andersson & J. Harris (Eds.), *Re-theorising the recognition of prior learning*. Leicester: NIACE.

- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405-450.
- SIAST. (2000). *Guide to prior learning assessment and recognition at SIAST*. Saskatoon: Saskatoon Inst. of Applied Science and Technology.
- Sireci, S. G. (2003). *Validity (General)*: SAGE Publications.
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher, 36*(8), 477-481.
- Starr-Glass, D. (2002). Metaphor and totem: exploring and evaluating prior experiential learning. *Assessment & Evaluation in Higher Education, 27*(3), 221-231.
- Starr-Glass, D., & Schwartzbaum, A. (2003). A liminal space: challenges and opportunities in accreditation of prior learning in Judaic studies. *Assessment & Evaluation in Higher Education, 28*(2), 179-192.
- Stowell, M. (2004). Equity, justice and standards: assessment decision making in higher education. *Assessment & Evaluation in Higher Education, 29*(4), 495-510.
- Sundström, A. (2009). *Developing and validating self-report instruments. Assessing perceived driver competence*. Umeå University, Umeå.
- Sweygers, A., Soetewey, K., Meeus, W., Struyf, E., & Pieters, B. (2009). Portfolios for prior learning assessment: Caught between diversity and standardization. *The Journal of Continuing Higher Education, 57*(2), 92-103.
- Taylor, T. (1996). Learning from experience: Recognition of prior learning. *Asia-Pacific Journal of Teacher Education, 24*(3), 281-292.
- Tight, M. (1998). Lifelong learning: opportunity of compulsion? *British Journal of Educational Studies, 46*(3), 251-263.
- Trowler, P. (1996). Angels in marble? Accrediting prior experiential learning in higher education. *Studies in Higher Education 21*(1), 17-29.
- Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioural research: Importance and calculation. *Journal of Behavioural Education, 10*(4), 205-212.

- Wedman, I., Stoor, M., Carling, E., Djuvfeldt, G., Holmström, P., & Linder, J. (2007). *Validering av kunskaper och kompetens* (Report, Valideringsdelegationen). Gävle: Högskolan i Gävle.
- Whittaker, S., Whittaker, R., & Cleary, P. (2006). Understanding the transformative dimension of RPL. In P. A. J. Harris (Ed.), *Rethorising the recognition of prior learning*. Leicester: NIACE.
- Winter, R. (1994). Work-based learning and quality assurance in higher education. *Assessment & Evaluation in Higher Education*, 19(3), 247-257.