UvA-DARE (Digital Academic Repository)

Stimulating intercultural intellectual capabilities in intercultural communication: testing an innovative course design

Antonenko, T.A.

Publication date
2010

Citation for published version (APA):
Antonenko, T. A. (2010). *Stimulating intercultural intellectual capabilities in intercultural communication: testing an innovative course design*. [Thesis, fully internal, Universiteit van Amsterdam].

CHAPTER 3

DESIGNING AND TESTING INSTRUMENTS TO MEASURE INTERCULTURAL INTELLECTUAL CAPABILITY

An experiment is a question which science poses to Nature, and a
measurement is the recording of Nature's answer.
-- *Max Planck, German scientist*

Abstract

The purpose of this chapter is to select, adapt, design and test instruments to measure growth in Intercultural Intellectual Capabilities (IIC). Our model of intra- and inter-IIC growth has focused on critical cultural awareness, cultural stress tolerance and experimentation with the cultural self and related psychological traits like tolerance of ambiguity, emotional resilience, flexibility, creativity, openness, and cultural empathy. The selected measurement instruments must allow us to measure intra-IIC and inter-IIC through these clustered components and psychological traits. Because the higher order cognitive skills of the dimension of intellectual capabilities (dimension four) are at the core of IIC growth, and because a learner has to go through all four dimensions of intellectual development before IIC starts to grow, we decided to measure two complex cognitive skills separately: critical thinking and self-reflection. A fifth variable we aim to measure is intrinsic motivation. Intrinsic motivation gives an indication as to the way the course is perceived by students and therefore of importance for the attitude of the students towards the course and towards engaging themselves. It is a measure that may help to validate the course design.

Many measurement instruments to measure ICC have been developed and used in practical courses on IC over time (Fantini, 2006; Sinicrope, Norris and Watanabe, 2007). Out of these measures, we have selected and adapted the Intercultural Competence Adaptation Potential Scale (ICAPS-46), Intercultural Competence Assessment (INCA), Motivated Strategies for Learning Questionnaire (MSLQ), and Intrinsic Motivation Inventory Questionnaire (IMI) instruments to measure intra- and inter-IIC, critical thinking, and intrinsic motivation. We have used a combination of the MSLQ instrument and items created by ourselves to de-

velop a scale to measure self-reflection, which we name the Self-reflection instrument. By comparing the psychological traits each instrument is covering with the traits that need to be measured in order to measure intra- and inter-IIC, we confirmed theoretical validity of the instruments.

Following the instrument selection, we have tested the five measurement instruments on three different occasions regarding their reliability and internal validity. For each of the measurement instruments we have looked at ways to increase reliability and internal validity. For ICAPS, we reduced the number of items from 55 to 46, the coding of the open questions of the INCA questionnaire was changed from a 1 – 3 scale to a 1 – 5 scale, the number of items of the MSLQ-CT instrument was increased from five to eight. For the IMI instrument, we reduced the number of items first from 45 to 40 by removing the open-ended questions and two items on 'friends' that were not understood by the students, and later, we also removed the factor 'pressure' because in our approach pressure is a positive factor for creating intrapersonal disequilibria, while for the instrument, pressure was viewed as negative for motivation. We combined part of the MSLQ scale on self-regulation with our own items to create an instrument to measure self-reflection.

The ICAPS questionnaire (adapted from Matsumoto et al., 2001) was tested for reliability and it turned out to be between .75 and .90, which is high. We decided to use the instrument as a whole, and not its individual scales. The INCA instrument (2007) showed an overall level of internal consistency of between .69 and .96. We decided to use only the aggregate INCA instrument. Pearson correlations between the two instruments show a highest correlation of .24, which is sufficiently low to conclude that they measure something distinctly different. Our MSLQ-CT instrument for critical thinking (adapted from Pintrich et al., 1991) was tested and found reliable (reliability between .72 and .86). The Self-reflection instrument was also found highly reliable (reliability between .74 and .81), indicating a high degree of internal consistency. Correlation analysis shows that critical thinking and self-reflection had a correlation coefficient of between .66 and .75 (with one reading spiking to .81). This is a rather high level of correlation, but still there is a significant part of unexplained variance to warrant treating the two components separately. Our IMI instrument proved to be reliable. Reliability tests showed values between .75 and .97.

Having tested for all measurement instruments at three different moments in time, we find that – having made some changes to the instruments at various points – reliability and validity of the measurement instruments have been confirmed. We are therefore confident that ICAPS-46 instrument, the adapted INCA instrument, the MSLQ-CT and Self-reflection instruments, as well as the IMI instrument – in their revised forms – can be used to adequately measure the effect of our experimental course on inter- and intra-IIC growth, growth of critical thinking and self-reflection skills, and changes in intrinsic motivation.

Key words: measurement instruments, intercultural intellectual capability, intercultural communicative competence, critical thinking, self-reflection, validity, reliability

1.   FROM OUR MODEL OF IIC TO COMPONENTS OF ICC
AND PSYCHOLOGICAL TRAITS

1.1 Intercultural intellectual capabilities (IIC)

In chapter 2, we have described our model for growth in IIC. According to experiential psychology, experiences can trigger development on the perceptual, affective and symbolic levels of thinking, provided that the person who lives these experiences self-initiates reflection and/or is committed to active experimentation with the self and the new insights gained from this reflective intrapersonal dialogue. Intercultural experiences matter. Specific experiences – through intrapersonal contemplation and/or through interpersonal encounters – cause disequilibria in the minds of learners and set a cognitive process of adaptation and adjustment going. This process runs through four dimensions of the human intellect: the cognitive dimension, the meta-cognitive dimension, the intentional dimension and the intellectual capability dimension. Growth of these dimensions leads to growth in our model of intercultural intellectual capabilities (IIC).

In the dimension of intellectual capabilities (the fourth dimension), complex cognitive processes take place. Two of the most important complex cognitive processes are critical thinking and self-reflection (see chapter 2). Learners explore personal traits and values that – because of created mental disequilibria following intercultural cognitive experiences – can adapt or modify to understand and accept the differences of other cultures through a process of critical thinking. Critical thinking is the persistent and intentional effort to examine any cultural belief or experience in light of new evidence that does or does not support it. Self-reflection is analytical introspection whereby through continuous reconstruction of new intercultural information the self is checked, tested and – if needed – adapted and adjusted. Critical thinking and self-reflection are part of IIC in that an adequate choice of behaviour that has been thought over and reflected upon is part of intellectual capability. Moreover, complex cognitive skills could be employed to reset one's earlier semantic structures and ways of encoding information. In essence, IIC focuses on the intrapersonal process underlying the acquisition of ICC, essentially opening up the 'black box' of how learners acquire intercultural capabilities. This implies that some of the components of ICC that are related to inner cognitive developmental processes are also coming back in our model of IIC.

Comparing our model to other models of intercultural maturation (Matsumoto et al., 2001, 2005; King and Baxter Magolda, 2005; INCA, 2007), we find that these other models specify different dimensions for intercultural growth and maturation, like intrapersonal developments, interpersonal developments and cognitive developments. From these models, we observe that for growth in IIC, complex intellectual and psychological processes take place in the mind of the learner, like critical thinking, problem solving, and self-reflection. This implies that advanced intellectual skills are needed. Our model of IIC encompasses these previous models and in addition lays bare that development in the intellect is subject to change at different

levels of an individual's intercultural intelligence, i.e. the four dimensions of the human intellect mentioned above.

1.2 Linking our model of IIC to components of ICC

Inside each of the four dimensions of the human intellect a process of new information, combined with new processing skills implies potential for growth in IIC. More specifically, in order to develop IIC, we are looking for components of ICC that are most affected during the four-dimension growth of intercultural maturity. Many components of ICC have been identified in the literature up to now – for an overview see Spitzberg (1993), Fantini (2006), Risager (2006; 2007). However, as explained in chapter 2, even if other components of ICC have been laid bare before, we focus on developing critical cultural awareness, experimenting with the cultural self, and cultural stress tolerance as the main components of ICC. We do this, because these are the important components that matter both in intercultural competence and in intellectual learning processes at the same time. These three components of ICC are the three sub-attainment targets for the course. They also come back in the various dimensions of the human intellect in our model of IIC.

Cultural stress tolerance is developed in the meta-cognitive dimension, when meta-cognitive experiences lead to a conscious choice to control cultural stress and uncertainty. In the intellectual capabilities dimension, divergent capabilities include finding creative and new solutions to intercultural challenges, which in turn increases tolerance for cultural stress.

Critical cultural awareness, a critical look upon new knowledge of another culture, starts to grow in the meta-cognitive dimension, when intercultural knowledge is viewed meta-cognitively and when openness to other cultures develops. Also in the intentional dimension, critical cultural awareness is encouraged, when intentional experiences help to create new views, frames of mind and preferences. Critical cultural awareness also follows directly from the intellectual capability process development and from new perceptions to intercultural knowledge.

Willingness to experiment with the cultural self is found in the meta-cognitive dimension where openness to otherness and openness to new intercultural information is developed as well as where experimentation with different levels of control of intercultural information can be envisaged. It is also the result of the fourth dimension where IC participants that have been open to changes in their basic architecture, engage in complex thinking processes, like critical thinking and self-reflection. Divergent and convergent capabilities lead the learners to experiment with their own cultural selves.

1.3 Psychological traits considered favourable to ICC

We have identified various psychological traits from the literature that relate to inner cognitive developmental processes and that are mentioned frequently in academic literature as favourable to growth of ICC. These traits centre around a person's ability to adjust and adapt. According to Byram (1989) and Kim (1994), ICC is com-

prised of (socio)linguistic, cognitive and operational (behavioural), and affective (capacity to deal with stress) dimensions. Matsumoto et al. (2001) added emotional regulation management (or emotional resilience) in response to intercultural experiences as a pre-requisite for intercultural adaptation and adjustment, as well as other psychological constructs like openness, behavioural flexibility, creativity, tolerance for ambiguity, emotional resilience and cultural empathy. These have been discussed in chapter 2 (Lundstedt, 1963; Brein and David, 1971; Ruben, 1976; Furnham and Bochner, 1982; Furnham, 1986, Furnham,1987; Brislin and Yoshida, 1994; Kealey, 1996; Byram, 1997; Matsumoto et al., 2001; Sercu, 2002; Riemer, 2003; Williams, 2005; Matsumoto et al, 2005; Abarbanel, 2009; Osland, Bird and Gundersen, 2010).

1.4 Linking components of ICC to psychological traits favourable to ICC

The psychological traits overlap with and are closely linked to the three sub-attainment targets for the course on IIC.

Cultural stress tolerance is closely linked to behavioural flexibility, because the more flexible an IC participant, the lower cultural stress levels and the higher stress tolerance will be. It is also linked to - and even partially overlaps with – tolerance for ambiguity, because the higher the levels of tolerance of an IC participant, the lower the levels of cultural stress tolerance. Emotional resilience matters a great deal to keep cultural stress in check – not giving in to negative emotions when placed in uncomfortable situations increases cultural stress tolerance. Creativity matters for cultural stress tolerance, because creative learners are more flexible and dynamic in finding solutions to cultural stress, thus developing more tolerance against it. Finally, cultural stress tolerance is expected to be lower, when a learner is more culturally empathic, because when it is easier to stand in someone else's shoes and understand their ideas and viewpoints, stress tolerance will be much higher as the IC participant is more comfortable with differing views.

Critical cultural awareness requires openness, cultural empathy, and behavioural flexibility. When a learner is open to new experiences, new information and is open to otherness, awareness of different cultures and differences is much easier to obtain. Critical awareness involves reflecting upon different information and engaging in higher level cognitive activities like critical thinking and self-reflection. An IC participant with a high degree of cultural empathy will find it much easier to be critically culturally aware because it is relatively easier to get underneath someone else's skin and understand where other points of view, behaviour, and reactions come from. Behavioural flexibility matters – especially in the intentional dimension – because it is a driver for intentionally searching for and engaging in dialogues with others and being open to new intercultural frames of mind, views and preferences.

Willingness to experiment with the cultural self relates to a large number of the above-mentioned psychological traits. A person who wants to experiment will have to be open and flexible in terms of behaviour because experimentation may require different behaviours and an open mind to the process throughout the experiment. It also relates to emotional resilience and creativity. Only an emotionally resilient IC participant will truly be able to experiment with the cultural self, possibly facing

negative emotional experiences. It also helps to be creative so experimentation can go in different hitherto unexplored directions whereby creative solutions must be found, i.e. divergent capabilities are employed.

1.5 Intra-personal and inter-personal growth

Growth of IIC is in essence an intra-personal process stimulated by interpersonal experiences. The learner goes through the four intra-personal intellectual dimensions of the model – dimensions that also have interpersonal elements – in order to become a more mature intercultural communicator, and is subject to two main forces. First of all, the force of individual experiential learning, whereby new information and experiences cause disequilibria at the cognitive level, setting a process of intercultural maturity in motion. In addition to the intra-personal process, interpersonal experiences and information – coming from interaction with others – also provide information to the learner adding more disequilibria, strengthening the intra-personal growth process. The second force is the interpersonal process of a dialogue that is in essence the exchange of opinions between actors that at the same time develop intra-personally, sharing with each other these experiences and this growth. At the end, the intra-personal growth and maturity of a person – following from going through cognitive, meta-cognitive and intentional developments multiple times – is seen in the way interactions with others are carried out, i.e. at a more mature intercultural level.

The two dimensions of IIC, we named intra-IIC – relating to the intrapersonal dimension of IIC – and inter-IIC – relating to the interpersonal dimension of IIC. The division between them is relative as they are interrelated. Intra-IIC focuses on how to deal with emotions and internal cognitive processes while inter-IIC focuses more on how to deal with situations and others, and both of these feed the intra-personal IIC growth process.

Experimentation with the cultural self is mostly an intrapersonal experience or process because, requiring openness to otherness and acceptance of possible changes in the basic cognitive architecture. It is, however, also based on and can draw from interpersonal communication and interpersonal exchange of information and exchange of cultural identities.

Cultural stress tolerance reflects upon intrapersonal experiences and learning how to deal with cultural stress, how to control negative emotions and filter intercultural information. Intrapersonally, cultural stress tolerance is also about allowing divergent capabilities lead to new and creative solutions to intercultural problems. It is also grown by learners when engaging in a dialogue or by being confronted with uncomfortable cultural differences in other IC participants, i.e. by interpersonal experiences.

In terms of the nature of the process, experimentation with the cultural self and cultural stress tolerance are more intrapersonal than interpersonal experiences, and that is why we refer them more to intra-IIC than to inter-IIC.

Critical cultural awareness – even though linked to, for example, tolerance for ambiguity as well – is mostly based on intercultural situations and confrontations,

intercultural behaviour rooted in intrapersonal intercultural development, and exchange of information.

The above sections are summarised in Table 3.1. The links between our model of IIC, (especially the four specific dimensions), the detailed explanations of those dimensions, and the components of ICC that matter both in intercultural competence and in intellectual learning processes at the same time (i.e. critical cultural awareness, cultural stress tolerance and willingness to experiment with the cultural self) are presented in columns (1), (2) and (4). The cognitive dimension is clearly the starting point, while in the other three dimensions, the components of ICC grow.
Column (3) shows whether a process is mainly intra- or interpersonal. From this column we observe that in each dimension of the model, both intra- and interpersonal processes run and that – when combining column (3) and (5), several of the psychological traits have both an intra- and an interpersonal dimension. The links between these three components of ICC and related psychological traits that affect intercultural competence favourably, are summarised in columns (4) and (5).

From our theoretical model of IIC growth and the link to components and psychological traits of ICC presented in Table 3.1, we conclude that we have to measure both intrapersonal and interpersonal development of the components of ICC, i.e. critical cultural awareness, cultural stress tolerance and willingness to experiment with the cultural self. Since a learner has to go through all four dimensions of our model, eventually the one on growth in intellectual capabilities, we also want to look at whether complex higher order cognitive processes take place. Therefore, we choose to investigate separately whether the course on IIC – next to stimulating growth of intra- and inter-IIC – includes stimulating complex cognitive processes like critical thinking and self-reflection.

Table 3.1 Linking the model of IIC to ICC components and psychological traits

Dimensions of model of IIC	Explanations of dimensions of IIC	Intra- or interpersonal experiences	Components of ICC relevant for IIC	Psychological traits favourable to ICC growth
(1)	*(2)*	*(3)*	*(4)*	*(5)*
Cognitive dimension	New intercultural semantic structures	Intra		Openness
	Motivation to create new intercultural meaning	Inter		Openness
	New intercultural cognitive schemes	Intra		Openness
	New intercultural information coding ways	Intra		Openness
Meta-cognitive dimension	Motivation to control all information against an intercultural filter	Intra	Cultural stress tolerance	Behavioural flexibility, tolerance for ambiguity, emotional resilience
	Openness to otherness and willingness to dialogue and reconsider acquired schemata	Intra	Critical cultural awareness Experimentation with the cultural self	Tolerance for ambiguity, emotional resilience, cultural empathy, openness, creativity
	Motivation to control cultural stress and tolerate cultural stress	Intra	Cultural stress tolerance	Behavioural flexibility, tolerance for ambiguity, emotional resilience, cultural

Table 3.1 Linking the model of IIC to ICC components and psychological traits

Dimensions of model of IIC	Explanations of dimensions of IIC	Intra- or interpersonal experiences	Components of ICC relevant for IIC	Psychological traits favourable to ICC growth
(1)	*(2)*	*(3)*	*(4)*	*(5)*
	and un-certainty		Experimentation with the cultural self	empathy
	Motivation to add more and classify new knowledge and information from others	Inter	Critical cultural awareness Experimentation with the cultural self	Behavioural flexibility, tolerance for ambiguity, cultural empathy, openness
Intentional dimension	Intentional search for new intercultural preferences	Intra	Critical cultural awareness	Behavioural flexibility, openness
	Motivation to gather more different information in dialogue with others	Inter	Critical cultural awareness	Behavioural flexibility, tolerance for ambiguity, cultural empathy, openness
	Intentional engagement with other intercultural frames of mind and views	Intra	Critical cultural awareness	Behavioural flexibility, cultural empathy, openness
Intellectual capabilities dimension	Skills to analyse intercultural experiences in convergent ways	Intra and inter	Critical cultural awareness	Openness
	Skills to analyse intercultural	Intra and inter	Cultural stress toler-	Behavioural flexibility, tolerance for am-

Table 3.1 Linking the model of IIC to ICC components and psychological traits

Dimensions of model of IIC	*Explanations of dimensions of IIC*	*Intra- or interpersonal experiences*	*Components of ICC relevant for IIC*	*Psychological traits favourable to ICC growth*
(1)	*(2)*	*(3)*	*(4)*	*(5)*
	experiences in divergent ways		ance	biguity, emotional resilience, openness, creativity
	Skill to be critical of one's own thinking and reflect upon this	Intra and inter	Experimentation with the cultural self	Behavioural flexibility, tolerance for ambiguity, cultural empathy, creativity
	Developed learning ability	Intra and inter	Critical cultural awareness Cultural stress tolerance	Behavioural flexibility, tolerance for ambiguity, emotional resilience, cultural empathy, openness, creativity
	New intercultural know-ledge perception styles	Intra and inter	Critical cultural awareness	Behavioural flexibility, cultural empathy, openness

2.   SELECTION OF MEASUREMENT INSTRUMENTS

2.1 Review of existing measurement instruments

There is quite a large variety of instruments available that have been elaborated to measure components of ICC (Fantini, 2006; and Sinicrope, Norris and Watanabe, 2007), as well as intellectual development (Gottfredson, 1997). They all provide extensive overviews. These instruments cover various foci. Some are aimed at specific countries (Pruegger and Rogers, 1994; Black, 1988) or specific purposes like work abroad, doing management tasks abroad or team processes (Casse, 1982; Kealey, 1988), or linguistic competences (ACTFL, 1982; Ingram and Wylie, 1984, 1997).

The Cultural Competence Self-Assessment Questionnaire (CCSAQ) by Mason (1995) focuses on working with children with disabilities and their families. Corbitt (1998) developed the GAP Test: Global Awareness Profile, that focuses on how much world knowledge a person has concerning items like politics, economics, geography, and culture.

Other instruments test (part of) the ICC competences relevant for intrapersonal and interpersonal development in IIC. The Behavioral Assessment Scale for Intercultural Communication (BASIC) of Koester and Olebe (1988) and Olebe and Koester (1989) looks at assessing behavioural traits from an interpersonal perspective, and so do the Multicultural Counseling Inventory or MCI (Sodowsky, Taffe, Gutkin, and Wise, 1994), and The Multicultural Counseling Awareness Scale or MCAS (Ponterotto et al. (1991). The Beck Depression Inventory (BDI) links to ICC components in that it views depression as sustained by intrusive negative cognitions (Beck, 1996), i.e. a lack of cultural stress tolerance and ability to handle uncertainty. BEVI – Beliefs, Events, and Values Inventory – by Shealy (2004) looks at developmental, affective and attributional processes and explains processes by which beliefs, values and 'world views' are acquired and maintained. The focus is on how 'open' people are to various transformational experiences.

With 18 scales grouped into four clusters, the California Personality Inventory (CPI) (Gough, 1987) studies: (1) measures of poise, ascendancy, self-assurance, and interpersonal adequacy; (2) measures of socialisation, responsibility, intrapersonal values, and character; (3) measures of achievement potential and intellectual efficiency; (4) measures of intellectual modes and interest modes. The Cross-cultural Adaptability Inventory (CCAI) of Kelley and Meyers (1988) and of Grove and Associates measures emotional resilience, flexibility and openness, perceptual acuity, and personal autonomy while the Intercultural Development Inventory (IDI) focuses more strongly on measuring individual's intercultural development ranging from ethnocentrism to ethnorelativism (Bennett, 1991, 1993, 1994; Bennett and Hammer, 1998). The latter index was tested in corporate and educational settings.

The Intercultural Sensitivity Inventory (ICSI) by Bhawuk and Brislin (1992) measures intercultural sensitivity based on concepts of individualism and collectivism of Hofstede (1980, 1991).

When investigating empathy as a psychological trait, the Model of Intercultural Communication Competence developed by Arasaratnam (2004) and Arasaratnam and Doerfel (2005) is a good choice since it focuses mainly on the link between ICC and empathy. Studying personality measures – that matter for the way intrapersonal views on interaction manifest themselves – the California F Scale looks at different traits associated with the authoritarian personality like conventionalism, authoritarian aggression, anti-intraception and power and toughness (Hohn, 1972). The California Critical Thinking Disposition Inventory (CCTDI) (Facione, Sanchez, Facione, Gainen, 1995) is an important tool to survey dispositional aspects of critical thinking. It contains 75 statements about beliefs, values, attitudes and intentions, and focuses on measuring the willingness to think critically. The Eysenck personality inventory or EPI (Eysenck and Eysenck, 1975) covers various psychological traits relevant for ICC when it looks at personality as two biologically-based categories of temperament: extraversion vs. introversion (different optimal levels of arousal resulting in intercultural performance) and neuroticism vs. stability (high of low levels of stress tolerance).

Byram (1997), Byram, Nichols and Stevens (2001), Kuhlmann and Stahl (1998), and Muller-Jacquier (2000) were in charge of the Intercultural Competence Assessment (INCA) project for the Leonardo da Vinci II programme of the EU (2007). They defined ICC and developed a framework and diagnostics tool for the assessment of ICC in terms of tolerance for ambiguity, behavioural flexibility, communicative awareness, knowledge discovery, respect for otherness, and empathy, with an emphasis on interpersonal engagement.

In order to measure personality structure and psychopathology to look at mental health, the Minnesota Multiphasic Personality Inventory (MMPI) was developed by Hathaway and McKinley (1940). It looks at issues like awareness of problems and vulnerabilities in learners, conflict, struggle, anger, and respect for society's rules, level of trust and sensitivity, odd thinking, level of excitability and people orientation. The Overseas Assignment Inventory (OAI) (Tucker, no date) analyses 14 attitudes and attributes related to successful cross-cultural adjustment and performance.

A broad instrument for looking at international competencies and how to be highly effective operators is provided by TCO International (Trickey, 2004). They include the competencies of openness, flexibility, personal autonomy, emotional resilience, perceptiveness, listening, orientation, transparency, cultural knowledge and influencing synergy in their test items. The State Trait Anxiety Inventory (STAI) instrument by Spielberger, Gorsuch, Lushene, Vagg and Jacobs (1983) distinguishes between emotional and cognitive components in looking at intercultural competence.

An important index with respect to a learner's reactivity is the Interpersonal Reactivity Index (Davis, 1980; Davis, 1983) that was created to measure four factors related to empathy: the fantasy scale (imaginative empathy), perspective taking (tendency to take the psychological point of view of others, i.e. cognitive empathy), empathic concern (sympathy and concern for others – emotional empathy), and personal distress (feelings that get in the way of helping others).

Matsumoto et al. (2001) created the Intercultural Adjustment Potential Scale (ICAPS) based on social psychology of adjustment through an analysis of intercultural encounters and apply it to Japanese sojourners and immigrants to the US. This index builds on and incorporates various of the other testing instruments mentioned in this review (e.g. EPI, BDI, STAI, MMPI, and CPI). The components of ICC reflected in ICAPS are ability to regulate emotions, being free of over-attachment to previous ways of thinking, willingness to tolerate ambiguity, openness and flexibility, critical thinking, behavioural flexibility.

Finally, the NEO PI (Costa and McCrae, 1985) or – more recently – the NEO PI-R (Costa and McCrae, 1992) model is a 240-item measure, also called the Five Factor Model (Digman, 1990), looking at extraversion, agreeableness, conscientiousness, neuroticism (including inter alia six of its subordinate dimensions), and openness to experience.

In addition to the literature related to ICC and its components, there are also various measurement instruments directed more towards intelligence and intellectual development (Gottfredson, 1997; Gottfredson, 2004). While a lot of research has been done (Hogan, 1991; Schmidt, Ones and Hunter, 1992; Lubinski and Dawis, 1992), we focus on a few tests that have been developed and used to measure intelligence – or 'g' as intelligence is also called (Gottfredson, 1997). The US employment services have used the General Aptitude Test Battery (GATB) to measure aptitude of employees for jobs of different levels of complexity (Hunter, 1983, 1986). Components of the GATB are arithmetic reasoning, vocabulary, 3D-space, computations, name comparisons, tool matching and form matching – all testing cognitive skills and intelligence levels (US Department of Labor, 1970).

The Motivated Strategies for Learning Questionnaire (MSLQ) by Pintrich, Smith, Garcia and McKeachie (1991) also looks at cognitive components and shows high levels of predictive validity and reliability (Pintrich, Marx and Boyle, 1993). The testing instruments taps into: value (intrinsic and extrinsic goal orientation, task value), expectancy (control beliefs about learning, self-efficacy) and affect (test anxiety). The cognitive scale in Pintrich et al. (1991) includes among others elaboration and critical thinking and self-regulation.

Another test has been used by the US Armed Forces – the Armed Forces Qualifying Test (AFQT), focusing on testing intelligence and IQ related to 'trainability' of new recruits. The Wonderlic Personnel Test (WPT) is a 50-item intelligence test that many employers have used to screen job applicants, notably assessing the aptitude of these job applicants for learning and problem-solving in a range of occupations. It assesses person's abilities to learn, adapt, solve problems, and understand instructions. Validity and reliability of the instrument compare favourably to other adult intelligence tests (Matarazzo, 1972). The Wechsler Adult Intelligence Scale is another instrument used to look at intelligence, especially through language and vocabulary types of tests (WAIS; Wechsler, 1981). The Position Analysis Questionnaire (PAQ) also focuses on elements of work and the work context and includes almost 200 elements in categories like overall mental difficulty, communicating judgments, verbal aptitude, processing information, making decisions, and using various information sources (Harvey, 1991).

Finally, the National Adult Literacy Survey (NALS) is an assessment focused more on literacy (Kirsch, Jungeblut, Jenkins and Kolstad, 1993). It provides insights not only in literacy, but also in people's general ability regarding information-processing complexity. "Careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies" (Kirsch et al., 1993: 9). The scale contains five levels, from matching single pieces of information or performing simple operations (level 1) to multiple-feature matches, integration and synthesis of information from complex messages or documents, and use of multiple sequential operations (levels 4 and 5).

2.2 Selection of measurement instruments

We have chosen the variables that need to be measured, based on a review of components of ICC within our model of growth of IIC. That means, we need adequate testing instruments to measure: critical cultural awareness, cultural stress tolerance and experimentation with the cultural form of the self. We do so keeping in mind the two dimensions to IIC growth: the intrapersonal development of IIC (adjustment and adaptability traits), and interpersonal development of IIC (dialogue with others). In order to analyse whether learners in the course on IIC go through and get stimulated to use higher level complex cognitive processes (from the dimension of intellectual capabilities), ensuring growth of IIC, we also want to employ instruments that test learners' levels of critical thinking and self-reflection. Also, in order to test how much open and willing to engage the students are during the course, we test their levels of intrinsic motivation.

In selecting the instruments for our research, we were guided by three main decision criteria:
(1) Relevance: the measurement instruments must fit to the theoretical framework and measure the variables that need to be measured. That is, the instruments have to fit and reflect upon the three components of ICC (critical cultural awareness, cultural stress tolerance and willingness to experiment with the cultural self) and the intra- as well as interpersonal dimensions of IIC in the most accurate way.
(2) Predictive validity and reliability: the measurement instruments must have levels of predictive validity and reliability as high as possible, i.e. they test properly what they are supposed to test and different instruments do not test the same constructs.
(3) Feasibility: apart from validity, we are also bound by practicalities in the choice for the measurement instruments. They have to be easy to carry out, clear in form, and written in 'basic' English for the students to fill out and to be properly understood. Moreover, it is important that the instruments can be used several times during the experiment and are suitable for an academic environment. For that reason, we cannot choose to use charts or portfolios, nor internet-based instruments, so as not to be limited by technical constraints.

Intrapersonal dimension of IIC

For the intrapersonal dimension of IIC, we have pre-selected five instruments based on the first decision criterion: the intrapersonal approach to development they measure and the fact that they cover most of the ICC components (and related psychological traits) required by the model:

1. The Cross Cultural Adaptability Inventory (CCAI) instrument by Kelly and Meyers (1988) / a commercial project by Grove and Associates;[7]
2. The Bennett (1991, 1993, 1994) and Bennett and Hammer's (1998) Intercultural Development Inventory (IDI) questionnaire;
3. The Intercultural Adaptation Potential Scale questionnaire (ICAPS) by Matsumoto et al. (2001);
4. The commercial TCO international competencies instrument (ICI) (Trickey, 2004);
5. The Wonderlic Personnel Test (WPT) (Wonderlic Personnel Test, 1992).

Table 3.2 Measurement instrument selection for the intrapersonal dimension of IIC

Name instrument	Criterion 1: Relevance (1)	Criterion 2: Predictive validity and reliability (2)	Criterion 3: Feasibility (3)	Total number of points (4)
1. CCAI (Kelly and Meyers, 1988) / Grove and Associates	2 + 2 + 3 = 7	2 (Kelly and Meyers, 1992; Davis and Finney, 2006)	4	13
2. IDI (Bennett 1991; Bennett and Hammer, 1998)	3 + 2 + 3 = 8	4 (Hammer et al., 2003; Paige, Jacobs-Cassuto, Yershova, DeJaeghere, 2003)	3	15
3. ICAPS (Matsumoto et al., 2001)	4 + 2 + 4 = 10	4 (Matsumoto et al., 2001, 2005)	5	19
4. ICI (TCO international, Trickey, 2004)	2 + 3 + 3 = 8	NA	3	11
5. WPT (WPT, 1992)	2 + 2 + 2 = 6	2 (WPT, 1992)	4	14

---

[7] *Available at: www.grovewell.com.*

Looking at the other three criteria combined, Table 3.2 shows how we apply the three decision criteria. For criterion 1 – column (1) – we gave one point for each dimensions of IIC (four maximum), component of ICC (three maximum) and related psychological trait (five maximum) in the model that is covered by the measurement instrument. This column is further substantiated in Figure 3.1 where for the chosen measurement instrument, we specify in detail what components of ICC and what psychological traits that facilitate growth in IIC are measured.

For criterion 2 – column (2) – we present the tested levels of predictive validity and reliability and rank the instruments from 4 (highest level) to 1 (lowest level). In this column, we also present the source article where these test results can be found. For criterion 3 – column (3) – we graded the feasibility by giving one point if easy to carry out, one point if clear in form, one point if 'rewritten' to basic English, one point if the instrument can be used repeatedly, and one point if the instrument is suitable for an academic environment. This makes a possible total of 5 points for criterion 3. Column (4) shows the total points for each measurement instrument.

In column (1), ICAPS scores ten and ICI six points. ICAPS measures all four dimensions of IIC (4), experimentation with the cultural self and cultural stress tolerance (2), while the ICI measures the cognitive, meta-cognitive dimensions (2), and all three components of ICC (3). However, as specific psychological traits, emotional regulation, openness, flexibility and creativity (4) are measured by ICAPS while the ICI measures openness, flexibility and emotional resilience (3). Testing of reliability and validity, we found that both the IDI and ICAPS instruments had been tested extensively and that both also showed high levels of reliability and validity (Hammer, Bennett and Wiseman, 2003; Paige, Jacobs-Cassuto, Yershova, De-Jaeghere, 2003; Matsumoto et al., 2001, Matsumoto et al., 2005). We therefore have given equal points to both instruments for this criterion. For the third decision criterion, the ICAPS instrument is easiest to administer (repeatedly) and least cost, especially compared to the IDI and ICI instruments and to a lesser extent to the WPT instrument.

From Table 3.2 it becomes clear that we have decided to use the Intercultural Adaptation Potential Scale (ICAPS) by Matsumoto et al. (2001). This instrument focuses on intrapersonal development and contains several of the elements that are important in our model for IIC development: tolerance for ambiguity (cultural stress tolerance of dimension 2 in our model of IIC), openness and flexibility (divergent capabilities – dimension 4) and openness to intercultural information in dimension 2 of our model of IIC), and emotional regulation (applying a cultural filter as in dimension 2 of our model on IIC). Also the questionnaire contains questions to look deeper into the set of personal position and beliefs in relation to intercultural change.

It focuses on psychological skills that are considered to be important in managing intercultural conflict and stress that are inevitable during a sojourn (Matsumoto, 1993). The validity and reliability of the ICAPS instrument have been tested by Matsumoto et al. (2001, 2005) and found more than satisfactory. Also other authors have used the ICAPS instrument and found its reliability and validity to be above standard (Savicki, Downing-Burnette, Heller, Binder, Suntinger, 2004). Another strong point of ICAPS compared to the other measurement instruments is the way the instrument has been constructed, "using earlier personality inventories that have

assessed psychological constructs like emotion regulation, critical thinking, openness and flexibility, interpersonal security, emotional commitment to traditional ways of thinking, tolerance of ambiguity, and empathy" (Matsumoto et al, 2005: 21-22). These personality inventories include the EPI, BDI, STAI, BSRI (Bem Sex Role Inventory in Bem, 1974; and Bem and Watson, 1976), MMPI (Tellegen, Ben-Porath, McNulty, Arbisi, Graham, and Kaemmer, 2003), CPI, NEO-PI-R, BFI, California F-Scale and IRI.[8] This has resulted in an instrument of 193 items that have then been re-arranged to yield the ICAPS-55 (55 items) instrument (Matsumoto et al., 2005). Moreover, the ICAPS questionnaire is readily accessible, unlike the IDI that is more costly and requires following a preparatory course first, before getting access to the questionnaires.

Interpersonal dimension of IIC

For the interpersonal dimension of IIC, we have pre-selected two instruments based on the first decision criterion: interpersonal approach to development they measure and the fact that they cover several of the components of ICC (and psychological traits) sought after by the model:

1)   The Behavioral Assessment Scale for Intercultural Communication (BASIC) of Olebe and Koester (1989);
2)   The Intercultural Competence Assessment (INCA) project elaborated within the Leonardo da Vinci II framework (2007)[9];

Like with the intrapersonal dimension of IIC, intra-IIC, above, we apply all three decision criteria to make a final selection, using the same scale as in Table 3.2. Again, column (1) is further substantiated in Figure 3.1 where for the chosen measurement instrument, we specify in detail what dimensions of IIC, components of ICC and what psychological traits that facilitate growth in IIC are measured. Only column (2) – because we have pre-selected two measurement instruments – has a maximum of 2 (top rank). Table 3.3 summarises the findings.

In column (1), the relevance of both instruments is analysed. INCA looks at the cognitive and intellectual dimensions of IIC (2), critical cultural awareness as well as at cultural stress tolerance (2) and at (behavioural) flexibility, empathy, tolerance for ambiguity and respect for otherness (4). The BASIC instrument measures the same dimensions of IIC (2), critical cultural awareness and cultural stress tolerance (2) as well as empathy, tolerance for ambiguity, and display of respect (3). In column (2), we have awarded two points to the BASIC instrument as Koester and Olebe (1988) show reliability and validity values that are within accepted parameters for scale construction. INCA has not been extensively tested on validity and reliability and therefore receives no points for this decision criterion. Finally, for feasibility, the INCA instrument gets four points, while BASIC gets three. INCA – being developed in the public domain – is more readily accessible than BASIC, though both are easy to administer and framed in accessible English language.

---

[8] *See the literature review of measurement instruments for the full names of these inventories.*
[9] *Available at* http://www.incaproject.com.

Table 3.3 Measurement instrument selection for the interpersonal dimension of IIC

Name instrument	Criterion 1: Relevance	Criterion 2: Predictive validity and reliability	Criterion 3: Feasibility	Total number of points (highest score)
	(1)	(2)	(3)	(4)
1. BASIC (Olebe and Koester, 1989)	2 + 2 + 3 = 7	1 (Olebe and Koester, 1989)	3	11
2. INCA (Leonardo da Vinci II framework, 2007)	2 + 2 + 4 = 8	NA	4	12

For measuring the interpersonal dimension of IIC, we have therefore chosen (see Table 3.3) for the Intercultural Competence Assessment (INCA, 2007) elaborated within the Leonardo da Vinci II project. This instrument is based on interpersonal dialogue, but also contains questions that test the development of critical cultural awareness (dimensions 2, 3 and 4 of the model of IIC) and covers tolerance for ambiguity (dimensions 2 and 4 (control of cultural stress and divergent capabilities) in our model of IIC), relatedness to otherness (dimension 3 through intercultural preferences and dimension openness to otherness in dimension 2 in our model of IIC) and intercultural communicative awareness (critical cultural awareness in dimensions 2, 3 and 4 in our IIC model). Moreover, INCA is a questionnaire with open questions – which is what we have been looking for, because it reduces bias in the answers from the IC participants (i.e. giving socially accepted answers), because they do not know how the answers are assessed (Sinicrope, Norris and Watanabe, 2007), even though this makes the work more laborious.

Critical thinking and self-reflection
Critical thinking and self-reflection are two specific and important complex cognitive processes that stimulate intrapersonal intellectual capabilities. Critical thinking and self-reflection are both outputs of the model on IIC (i.e. present in divergent thinking skills and knowledge perception styles in dimension 4 of our model) as well as two of its drivers. Critical thinking skills are needed to keep the acquisition process of IIC going, like, for example, in the intrapersonal response to created disequilibria stemming from new intercultural information and experiences (dimension 1 of our model on IIC).

For critical thinking, we have pre-selected two instruments based on the fact that they cover different critical thinking items and are both frequently used to measure critical thinking:
1) The MSLQ inventory by Pintrich et al. (1991);
2) The California Critical Thinking Disposition Inventory (CCTDI) – (Facione et al., 1995);

The results of the selection are presented in Table 3.4. Both are measuring critical thinking skills (decision criterion 1), which is presented in column (1). We note that the CCTDI measures more components of critical thinking overall, but that the MSLQ measures those components that are most relevant for our model of IIC, i.e. it requires IC participants to focus on capabilities and progress in critical thinking. Both cover two dimensions of our model of IIC; the cognitive and intellectual capability dimensions.

Table 3.4 Measurement instrument selection for the interpersonal dimension of IIC

Name instrument	*Criterion 1: Relevance*	*Criterion 2: Predictive validity and reliability*	*Criterion 3: Feasibility*	*Total number of points (highest score)*
	(1)	(2)	(3)	(4)
1. MSLQ inventory (Pintrich et al. 1991)	2 + 1 = 3	2 (Pintrich et al., 1991)	5	10
2. CCTDI (Facione et al., 1995)	2 + 1 = 3	2 (Facione, 1990, 1992; Facione et al. 1995)	3	8

In column (2) we look at predictive validity and reliability of the instruments. The MSLQ instrument scores a Cronbach alpha of .80 (Pintrich et al, 1991) and is known as one of the most reliable instruments to measure critical thinking. However, also the CCTDI scores high on reliability and validity (Facione, 1990, 1992; Facione et al. 1995), has been tested extensively, and focuses on critical thinking skills specifically. We have chosen therefore to give both instruments equal points for this criterion. The feasibility of the measurement instrument for our research (decision criterion 3) is higher for the MSLQ than for the CCTDI for two reasons: first of all the measurement instrument is more easily (repeatedly) and more cheaply available than the CCTDI, and – second – the instrument is easier to carry out and takes less time for the students to fill out. That means that in total, we have opted for the critical

thinking items of the MSLQ instrument (Pintrich et al, 1991) rather than the CCTDI instrument.

For self-reflection, no specific measurement instrument was available in line with our research requirements. Research on self-reflection has been carried out by Gay and Kirkland (2003), Stronge (2002), Holtom and Kenworthy-U'Ren (2006), and Danielewicz (2001), who show that indeed, self-reflection is an important higher order complex cognitive skill that fits to the fourth dimension of our model of IIC. We therefore opted to adapt parts of the MSLQ questionnaire (Pintrich et al., 1991) – especially the subscale of self-regulation – and to develop item questions ourselves in order to create a reliable and valid measure for self-reflection (SR). The details of instrument construction will be explained further below.

Intrinsic Motivation

With intrinsic motivation being a prerequisite for learning, reflecting the attitude of the students towards the course, it is not directly – but definitely indirectly – linked to the success of the course. Students with higher levels of motivation should be more inclined to be open and engaged in the course, thus positively affecting their levels of IIC. This implies that intrinsic motivation is a variable that should be used to aid course design validation (chapter 4).

For intrinsic motivation, we have pre-selected two instruments based on the fact that they cover different items related to motivation and that they are both frequently used to measure motivation:
1)   The MSLQ inventory by Pintrich et al. (1991);
2)   The IMI questionnaire by Ryan (1982) and Ryan and Deci (1992);

The results of the selection are presented in Table 3.5. Both are measuring intrinsic motivation skills (decision criterion 1), but the IMI instrument is specifically focused on measuring intrinsic motivation and covers eight motivational components, while the MSLQ instrument measures three (and is more focused on learning strategies), which is presented in column (1).

In column (2) we look at predictive validity and reliability of the instruments. The MSLQ instrument scores a Cronbach alpha of .80 (Pintrich et al, 1991) and is known as one of the most reliable instruments to measure critical thinking. However, the IMI instrument by Ryan (1982) and Ryan and Deci (1992) has also been extensively tested (McAuley, Duncan and Tammen, 1989; Markland, 1999) and was reported with Cronbach alpha values of .90 and higher. We have chosen therefore to give the IMI instrument one point more for this criterion. The feasibility of the measurement instrument for our research (decision criterion 3) is equally high for the MSLQ and for the IMI instruments. They are both easily accessible and easy to administer. That means that in total, we have opted for the IMI instrument by Ryan and Deci (1992), rather than specific items of the MSLQ instrument (Pintrich et al., 1991).

Table 3.5 Measurement instrument selection for intrinsic motivation

Name instrument	Criterion 1: Relevance	Criterion 2: Predictive validity and reliability	Criterion 3: Feasibility	Total number of points (highest score)
	(1)	(2)	(3)	(4)
1. MSLQ inventory (Pintrich et al. 1991)	3	1 (Pintrich et al., 1991)	5	9
2. IMIQ (Ryan and Deci, 1992)	8	2 (Markland, 1999; McAuley, Duncan and Tammen, 1989; McAuley, Wraith and Duncan, 1991)	5	15

2.3 Linking the measurement instruments to the model of IIC and its components of ICC

We have opted for ICAPS, INCA, MSLQ, IMI and the SR instrument as the five measurement instruments for our research. The first three and fifth relate directly to the theoretical model and measure required outputs. The fourth, relates to student attitudes to the course, thus measuring one of the prerequisites for a successful course: student motivations. The ICAPS instrument intends to measure the intrapersonal dimension of IIC with its three components of ICC, and INCA the interpersonal dimension with those same components. MSLQ is chosen to measure critical thinking and – in an adapted and complemented form – also for self-reflection. The IMI instrument measures students' intrinsic motivation levels.

Turning back to the first decision criterion, relevance, Figure 3.1 shows in detail how the measurement instruments relate to the components of ICC and psychological traits that need to be measured. In column (1), the three main measurement instruments are presented.[10] Column (2) shows the psychological traits that are measured and specified for each of the measurement instruments. From the intra- and interpersonal dimensions of IIC in column (6) and the dimensions of intellectual development in column (5) follow – as described in detail in chapter 2 and summarised above – the three main components of ICC we focus on (column 4) as well as

---

[10] *Intrinsic motivation (IMI measurement) is not presented here, because it is not measuring a direct output required by the theoretical model. Rather it measures an indirect condition that ideally needs to be satisfied to create optimal conditions within which the course is to take place.*

the psychological traits considered favourable to developing intercultural competence (column 3). The essence of Figure 3.1 is the comparison between columns (2) and (3). It is clear that the psychological traits measured via the three chosen measurement instruments (column 2) match the psychological traits and ICC components (columns 3 and 4) that need to be measured in order to be able to measure growth in IIC.

The ICAPS instrument measures emotional regulation (matches emotional resilience), creativity (matches creativity), openness (matches openness) and flexibility (matches openness) as well as approximations of experimentation with the cultural self, cultural stress tolerance, and to a much lesser extent critical cultural awareness through its four factors.

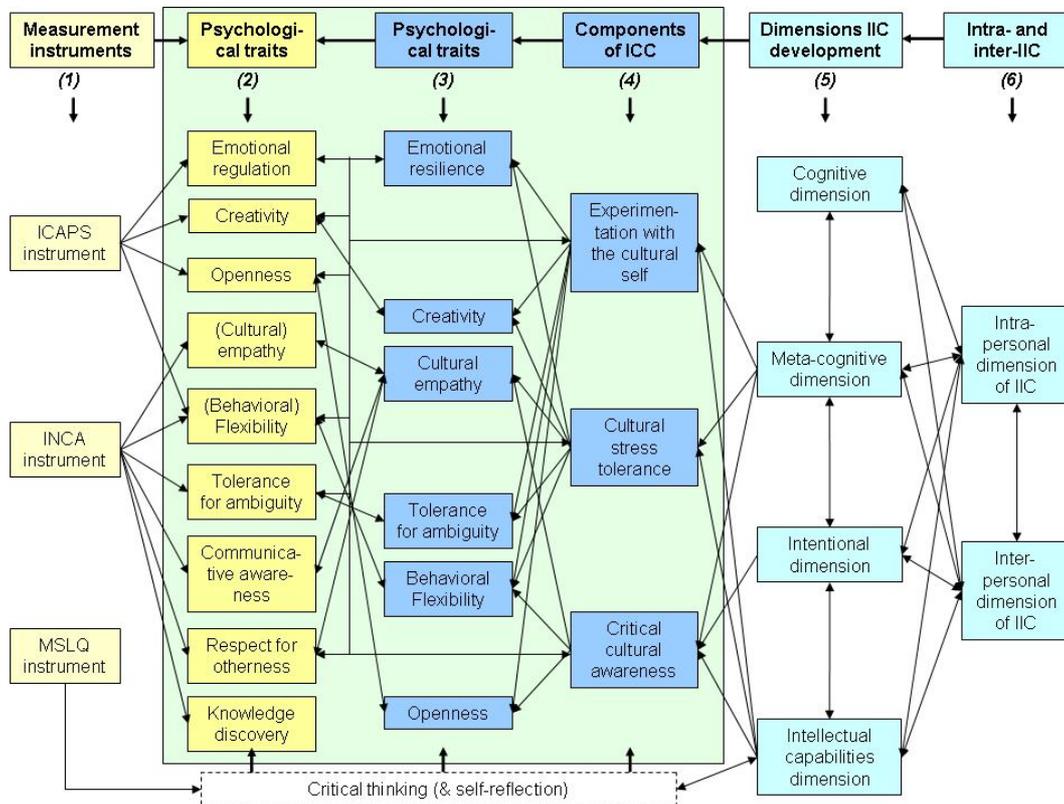*Figure 3.1 Linking measurement instruments to ICC components of our IIC model.*

The INCA instrument measures empathy (matches cultural empathy), behavioural flexibility (matches behavioural flexibility), tolerance for ambiguity (matches tolerance for ambiguity), communicative awareness (links to cultural empathy), respect for otherness (links to cultural empathy and respect for otherness), and knowledge

discovery (not directly linked) as well as to critical cultural awareness, and to a lesser extent to cultural stress tolerance. In addition, the MSLQ instrument measures critical thinking and (adapted) self-reflection which – according to Stronge (2002) and Danielewicz (2001) – act as complex cognitive skills that deepen and enhance the effects of the psychological traits and ICC components examined, thus enhancing Intercultural Intellectual Capabilities.

From Figure 3.1, we infer a few conclusions. First, that the ICAPS and INCA instruments – in theory – are well placed to measure the effects on our course on IC that is built on our model of stimulating development in IIC given their match in psychological traits. Second, the ICAPS and INCA instruments appear to overlap in part, but also to measure distinctly different psychological traits, reflecting the intra-IIC and inter-IIC dimensions. Third, the ICAPS and INCA instruments appear to overlap to some extent in the components of ICC they measure: both measure cultural stress tolerance (though from different angles), while ICAPS is also measuring experimentation with the cultural self, and INCA measures more strongly critical cultural awareness.

3. AIMS OF THIS CHAPTER

Having selected the measurement instruments ICAPS, INCA and MSLQ to measure intra-IIC, inter-IIC, critical thinking and self-reflection, and IMI to measure intrinsic motivation, the purpose of this chapter is to test these measurement instruments, adapt them if necessary to make them fit our research approach, and ensure sufficiently high levels of reliability and validity of the instruments. Thus the research question of this chapter is: 'Are the chosen measurement instruments reliable and valid? If not, to what extent can they be adapted to improve reliability and validity?'

4. TESTING THE MEASUREMENT INSTRUMENTS

4.1 The original measurement instruments

In this section, we explain in more detail how the measurement instruments were constructed, what types of questions were asked, and how we have tested and then – in case needed – adapted the measurement instruments for our research purpose.

ICAPS measurement instrument for measuring intra-IIC growth
The InterCultural Adjustment Potential Scale (ICAPS) has been developed by Matsumoto et al. (2001) and tested and further refined since (Matsumoto, Walbott & Scherer 2003). The purpose of ICAPS is to predict the intercultural adjustment potential among learners, i.e. predict the degree of difficulty a person would have in adjusting to a new culture and to describe the aspects of their mental processes development of intrapersonal capabilities. "ICAPS is intended to help individuals have the option of choosing to remain the same or to alter their patterns of thinking and interacting so that they may adjust more easily to a new culture" (LeRoux and Matsumoto, 2000: 322).

As in our model of IIC, the focus of ICAPS is on the underlying psychological components that matter for adaptation following cognitive disequilibria resulting in IIC growth. "The identification of several psychological variables as the keys to intercultural adaptation bypasses the problem of culture-specific knowledge, and views the potential for intercultural adjustment as a function of the psychological skills that individuals possess within them. In doing so, it does not rely on knowledge or attitudes […], or on the specific host culture in which [learners red.] reside, or on language proficiency." (Matsumoto et al., 2001: 486). Instead of reasoning from theory, ICAPS' psychological components are the result of empirical testing (factor analysis).

Matsumoto et al. (2001) examined questionnaires from a number of valid and reliable personality inventories, assessing the psychological constructs that are related to emotion regulation, critical thinking, openness/flexibility, interpersonal security, emotional commitment to traditional ways of thinking, tolerance of ambiguity, and empathy. These tests have been mentioned above. 193 Items were created at first. Difficulty of wording in English was checked and the questionnaires took into account the potentially limited English language capabilities of the (Japanese) IC participants. Then Matsumoto et al. carried out eight studies with ICAPS to demonstrate the reliability and validity of the instrument and to empirically select the most relevant items of the scale. We will summarise the steps and findings of these studies. For a detailed description of each study and the steps taken, we refer to Matsumoto et al. (2001). At first, ICAPS-193 was reduced to ICAPS-153 by eliminating items with low Pearson correlations. In the second study ICAPS-153 was further reduced to ICAPS-55 and a composite ICAPS score was created by averaging across the 55 items. Cronbach alpha for ICAPS-55 was found to be .78. ICAPS also turned out to be temporally reliable with Cronbach alpha values of .79 at time 1 and .84 at time 2 of an English-English and English-Japanese test-re-test, as well as convergently valid. Further testing revealed that ICAPS-55 was also constructed validly, and has predictive validity of intercultural adjustment – using multiple methods. Savicki et al. (2004) have also tested the validity and reliability of ICAPS-55 and confirm the findings of Matsumoto et al. (2001).

One of the keys, according to Matsumoto et al. (2001), to successful intercultural adjustment – a process that by nature of intercultural differences must be replete with struggles, frustrations and conflicts – is emotion regulation. When (negative) emotions are kept in check, clear thinking is possible and adjustment and adaptability are possible. This factor is measured with 14 items. Two items are for example: "I often worry about things that might go wrong" (factor loading of .571) and "I rarely feel anxious or fearful" (factor loading of -.720) (Ibid: 504). A second factor is openness to experience. This is particularly openness to ideas. It is measured in the instrument with 13 items. An item example is: "I have tried to write poetry" (factor loading of -.546) (Ibid: 504). The third factor within ICAPS-55 is flexibility to traditional ideas and social roles, named flexibility. This factor is measured with 14 items. "Sex education is a good thing" is an item on this scale with a factor loading of .723 (Ibid: 504). The final factor is coming from a desire for self-direction and freedom from arbitrary constraints. Matsumoto et al. (2001) term this creativity or

autonomy. This factor is measured with 14 items. An item example is "Spanking a child is the best way to teach them" (factor loading -.696) (Ibid: 504).

These four psychological factors are presented in Figure 3.1, in column (2) where they turn out to be measuring the psychological traits we are looking for in our research. The students are asked to express their opinions with respect to each of the statements by rating their answers on a 7-point Likert scale from 'strongly disagree' (1) to 'strongly agree' (7).

INCA measurement instrument for measuring inter-IIC growth
The INCA project, financed by Leonardo da Vinci II programme of the EU (2007) has developed a framework to address intercultural competence. During the project, six components were identified as important for ICC: tolerance for ambiguity, behavioural flexibility, communicative awareness, respect for otherness, empathy (coming from Kühlmann and Stahl (1998), and knowledge discovery (Byram, 1997). "An intercultural individual should be willing to engage in intercultural interaction ('motivation' – the affective component), should have the necessary skills and knowledge ('skills/knowledge' – the cognitive component), and should show those resources in his or her behaviour ('behaviour')" (Prechtl and Lund, 2007: 473).

The INCA measurement instrument is designed as follows: communicative awareness includes - 3 items, tolerance for ambiguity - 6 items, behavioural flexibility - 3 items, empathy - 5 items, knowledge discovery - 4 items and relatedness to otherness - 6 items. These factors are thought to be important for developing the interpersonal dimension of IIC (see Figure 3.1). Because the INCA questionnaire is based on open questions, we do not provide examples of each scale, but only of one, and further refer to www.incaproject.org where all information, questionnaires, assessors' guides, etc. is readily available. The level of empathy in a learner, for example, is tested through the following situation: "Your supervisor at work has asked you to socialise with a young foreign worker of the same age and gender as yourself, who speaks your language quite well. A. What topics of conversation would you choose? B. What kind of activities would you choose to do and why?" (INCA project, Exercise Intercultural Encounters: 4).

The INCA project produced two types of tests: cognitive/affective-oriented written exercises and behaviour-oriented group exercises. The main topic of questioning in the written exercises part of the INCA questionnaire are Western and Chinese cultures.[11] The student answers' to the questions referring to China, show what influences their knowledge and perceptions, travels, or more profound personal contacts with representatives of this culture. One of our expectations was that it would not be possible to acquire IIC only as a result of international encounters or travels. Following historical research by Otten (2003), Allport (1954) and Amir (1969), we argue in our model of IIC that intercultural encounters are maybe a necessary but not

---

[11] *This was especially interesting for Russian students because Russian culture (especially in the Asian part), apart from specific Russian cultural characteristics possesses both Western and Oriental views even though the degree to which these influences matter depends on the school and family education, personal experience, friends and travels.*

sufficient condition for acquiring IIC. For IIC to truly develop these encounters must start a process of intrapersonal development (measured by ICAPS).

The second part of the INCA questionnaire contains long scenarios united by one theme, Business trip to China, and requests students to show their reactions to the behaviour of the actors in the scenario and the problems they encounter, as well as the solutions they see in those intercultural situations. The instrument is available in pencil and paper format. Tests and assessment formats were generically constructed to be applicable for each country even though some were tested in a specific culture (e.g. China). Pilot exercises were conducted and modifications were made to all aspects of the INCA products, even though the pilot exercises only offered limited empirical data against which to validate the INCA grid. "From a diagnostic point of view, the INCA products can be seen as a stable, competence-based assessment…" (Prechtl and Lund, 2007: 473). However, more extensive testing is needed, in particular to further check the validity to intercultural competence and assessment to investigate wider applicability and usefulness of the INCA instrument (Armann, 2004).

MSLQ measurement instrument for measuring critical thinking and self-reflection
MSLQ is based on a cognitive approach to motivation scales and learning strategies. The motivation scales tap into expectancy, value and affect (Pintrich, 1988a, 1988b, 1989). The learning strategies are divided into cognitive (rehearsal, elaboration, organisation, critical thinking), meta-cognitive (planning, monitoring and regulating strategies), and resource management strategies (time and study environment, effort management, peer learning, help-seeking). As part of the cognitive strategy scale, MSLQ has a subscale on critical thinking, with two foci: applying previous knowledge to new situations and to make critical evaluations of new ideas.[12] MSLQ also has a subscale on self-regulation – defined as fine-tuning and continuous adjustment of one's cognitive activities reflecting on the self. In 1986, 1987 and 1988 large amounts of data were collected and the MSLQ was subjected to the tests on internal consistency, reliability and validity.

The Cronbach alpha coefficient for critical thinking was found to be .80 when MSLQ was tested by Pintrich et al (1991) and .80 in a sample of 380 Midwestern (US) college students (Pintrich et al., 1993). The critical thinking scale consisted of five items, like "I often find myself questioning things I hear or read in this course to decide if I find them convincing" (item 38) and "I try to play around with ideas of my own related to what I am learning in this course" (item 66) from Pintrich et al. (1991). Correlation among MSLQ scales (Pintrich et al., 1993: 812) show that for critical thinking no other factor measures the same. The highest correlations of the critical thinking scale are with intrinsic goal orientation (.58) and elaboration (.57)

---

[12] *These two foci are very relevant for our model on growth in IIC. Applying previous knowledge to new situations falls under 'combinative properties of the intellect' of convergent capabilities, 'originality' of divergent capabilities and learning ability – all components of the fourth dimension of our model on IIC growth, the intellectual capabilities dimension. Making critical evaluations of ideas relates to divergent capabilities (originality) and knowledge perception styles (cognitive styles).*

and the lowest (-.11) is with test anxiety. For the critical thinking subscale, the correlation with the final course grade was .15.

Self-reflection is not a separate scale in the MSLQ instrument, but some of the questions of the MSLQ instrument cover part of the definition of self-reflection as given by Stronge (2002) and Danielowicz (2001). Examples of items that also relate to self-reflection (adapted) are "When studying for this course, I set my own examples to help understand the concepts" (item 9), and "On each major concept from the study material I try to ask myself how I would react in the intercultural situations that illustrate it". Correlation among MSLQ scales (Pintrich et al., 1993: 812) show that for the subscale of self-regulation, no other factor measures the same.[13] However, we focus only later on self-reflection, when we combine some of the items from Pintrich et al (1991) with items created by ourselves.

Overall, MSLQ shows satisfactory reliability in terms of internal consistency, while also the subscales show a degree of predictive validity since they were related to academic performance in the directions expected.

IMI measurement instrument for measuring intrinsic motivation
The IMI questionnaire, developed by Ryan and Deci (1992), covers eight clusters of items: relatedness, interest or enjoyment, perceived competence, effort, pressure or tension, perceived choice, value, and motivation (Ryan and Deci, 1992). To Ryan and Deci (1992), the level of interest/enjoyment, for example, is closely related to their intrinsic motivation, because the higher the interest in the course (materials), the more motivated students will be. A question that was put to the students to check students' levels on intrinsic motivation are: "I think this course is very enjoyable". Pressure or tension, in Ryan and Deci (1992) are expected to be negatively correlated with intrinsic motivation. We believe – indeed trying to create tensions and pressure purposefully – that it is not the pressure and tension itself, but the lasting negative feelings about them, matter. An example item in this cluster is "I felt very tense while doing this activity". Effort is tested via items like "I tried very hard on this activity" or "It was important for me to do well at this task". Perceived competence is measured by items like "I think I am pretty good at this activity" and "I am satisfied with my performance at this task". A choice item is "I feel like it is my own choice to do this course". The scale on value (or usefulness) contains the item "I believe doing this activity could be beneficial to me". Finally, relatedness, that measures interpersonal contact and relationships, contains items like "I felt really distant to this person" and "It is likely that this person and I could become friends if we interacted a lot".

Cronbach alpha was found to be between .80 and .87 as tested by Markland (1999). McAuley, Wraith and Duncan (1991) found that scale reliability lay between .72 (pressure-tension) and .92 (interest-enjoyment), but that choice, with a Cronbach alpha of .54 stood out as the one negative outlier. Also an earlier test by McAuley,

---

[13] *We do not report the correlation findings here because they relate to a subscale of self-regulation which is only in part the same as our subscale on self-reflection.*

Duncan and Tammen (1989) found that reliability of the items was satisfactory (e.g. competence items showed a reliability of .84).

4.2 First (pre-test) redesign of the measurement instruments

Upon selecting the measurement instruments, we have studied them and made adaptations prior to the first testing moment (Y1) to make them more suitable for our research purpose.

Changes in the ICAPS-55 instrument

Before commencing with the test rounds, we changed the semantics in two of the items of the ICAPS-55 instrument to make the items more clear for the participating students in terms of level of English by replacing difficult words or rephrasing sentences. "I rarely feel anxious or fearful" (item 10) has been adapted to "I almost never feel worried, restless, or fearful", and "Being in tense emotional situations scares me" (item 4) has been adapted to "Being in stressful emotional situations scares me". We maintained the number of 55 items in ICAPS-55 at the start of the testing. The revised ICAPS-55 instrument is presented in Appendices A-2 and B-2.

Changes in the INCA instrument

The INCA instrument consists of four parts, which are: biographical and intercultural profile information, intercultural encounters, a business trip to China and a role play. We have reduced the elements from four to two for the first two testing moments (Y1 and Y2), focusing on the intercultural encounters and the business trip to China, because the extra biographical information was not needed in the testing phase and the testing operated under strict time limits because it took place in addition to the ongoing course of the students. During the full course, we added the biographical information, but not the role play. The biosheet was added to look for interaction effects with the IIC variables, and the role play was not added due to time and facility constraints. The role play was not included because of organisational challenges with video equipment that could not be arranged. Also for the INCA instrument, we changed some English language phrases to make them more easily understood for B1-C1 English proficiency students.[14] For example, at the scenario "Feeling Confused (1)", we have changed the sentence "Therefore you often feel out of your depth and confused" to "Therefore you feel uncomfortable and confused" (INCA project, Exercise Intercultural Encounters: 4). Because the role play was not included, the number of items changed from 27 to 19. Communicative awareness includes - 3 items, tolerance for ambiguity - 4 items, behavioural flexibility - 1 item, empathy - 3 items, knowledge discovery - 4 items, and relatedness to otherness - 4 items.

The original INCA instrument is presented in Appendices A-1 and A-4, and the revised one in Appendices B-1 and B-4. The INCA instrument distinguishes be-

---

[14] *Language levels vary between intermediate and pre-advanced and the classification and levels are based on the Common European Framework of Reference for Languages (2004).*

tween three levels of competence among learners for each of the scales (level one, basic competence; level two, intermediate competence; level three, full competence). Coding the student answers to the open questions, would then range between 1 – 3. Because we felt this would not create enough variance in the student answers, we have changed the coding range to 1 – 4, adding an intermediate competence level: basic, lower intermediate, higher intermediate, and full competence.[15]

Changes in the MSLQ instrument – critical thinking
The MSLQ instrument uses five items for measuring the subscale on critical thinking. First of all, we have adapted the semantics in three out of five items to reflect more closely that we refer to motivation related to teaching, i.e. a course on intercultural communication, and not to motivation in general. For example, "I often find myself questioning things I hear or read in this course to decide if I find them convincing" has been changed into "I often find myself questioning things I hear or read in this course on intercultural communication to decide if I find them convincing". Also some other textual aspects of the items were changed. For example "When a theory interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence" was changed to "When a theory, interpretation or practical example is presented in class or in the readings, I try to think about other examples from other sources to support what I hear and learn". In order to strengthen this subscale (since for our research we do not focus on the full instrument, but rather on this part), also in terms of validity and reliability, we enlarged the number of items from five to seven. The two items we added are: "When thinking about the intercultural difficulty, I try to decide what is the difference behind it" and "When thinking about the intercultural difficulty, I try to decide what I am supposed to expect from my partner in communication". The original and changed MSLQ-CT instrument are presented in Appendices A-5 and B-5.

Changes in the MSLQ instrument – self-reflection
The instrument designed to measure self-reflection contains 13 items. Seven of these items come in adapted form from the MSLQ instrument. Examples of adapted items are "During class time I often miss important points because I am thinking of other things" and "Before studying the material for each unit of the course, I try to visualize the problem in intercultural communication". The other six items are added based on the definition of Stronge (2002) to reflect analytical introspection to reconstruct incoming intercultural information continuously. These six questions include items like "If theories are difficult to understand, I try to compare them with my previous similar experiences to understand them better", and "I try to think of my own practical examples related to what I am learning in the course to be able to get more experience in dealing with intercultural differences". The Self-reflection instrument is presented in Appendix B-5.

---

[15] *Later on – as we will describe below, we decided to expand the scale even further to 1 – 5, with level 1 being basic competence and level 5 being fully competent.*

Changes in the IMI instrument

Before starting the first experiment, we reduced the original IMI questionnaire by five items: three items were removed because they were open questions and two items on relatedness were taken out because the Tomsk students did not understand and were overlapping with other items. This resulted in a 40 item instrument for the first experiment (see Appendix A-3). Also, because we conducted the instrument six times (two times at Y1, one time at Y2, and three times at Y3), whereby at moments 1, the future tense had to be used, and at moments 2 or three the past tense, we changed this as appropriate for the testing moments. Also we changed the word 'activities' for 'course' and 'person' for 'student / other students' to relate the questionnaire to our experiment that is about student-student learning. In the trial course, we continued revising the instrument by removing the factor 'pressure', reducing the number of items of the instrument further from 40 to 35. We did this because reliability of that factor was low (Cronbach alpha of .55) and because in our course on growth of IIC, discomfort, pressure and stress are important elements of the course that are not viewed negatively (as presented by Ryan and Deci, 1992) but as prerequisites to set intrapersonal cognitive processes going. The revised IMI questionnaire is presented in Appendix B-3 and used like that for the full course.

4.3 Testing design and methodology

We tested all three measurement instruments three times over a period of one year. In April 2008 (Y1), we organised a first session with 123 students to test the instruments. After this first test, we revised some of the instruments (e.g. changed scales, phrasing of items and scoring procedures) to improve their reliability and validity as well as increase student understanding of the items presented. In September 2008 (Y2), we tested them again while conducting the trial course on a small part of the full new course in IC, this time with 31 students participating. Finally, in the final experiment in March 2009 (Y3) (chapter 5), we established reliability and validity of the measurement instruments for 98 students at three measurement occasions. The results are reported in the remainder of this chapter and in Appendix C.

Participants

A total of 252 university students of different ages in their Linguistics and Intercultural Communication studies participated in the data collection to validate the instruments to be used for measuring the effect of the experimental IC course to stimulate IIC (chapter 5). The exact information about numbers of students, their ages and at what time they participated in the testing is provided in Table 3.6.

The Faculty of Foreign Languages at Tomsk State University (TSU) has an ethnically homogeneous population (all students are Russian with a few students from Kazakhstan). All students took an introductory course on Intercultural Communication (the experimental testing took place in April 2008 and the studies started in September 2007 for the second year students). The main criteria for participation in the testing session were sufficient levels of English to fill in the questionnaires and being a student from the Languages Department at TSU. Students would enroll on a

voluntary basis, but could obtain a certificate of successful participation from the University of Amsterdam and course credits at TSU in Tomsk.

Table 3.6. Student information on those participating in testing measurement instruments

	Testing instruments (Y1)	*Trial course run (Y2)*	*Full course run (Y3)*
Nr of students participating	123	31	98
Average age	17 – 22	19 – 23	17 – 25
% Males / % Females	13% / 87%	3% / 97%	9% / 91%
Language levels*	B1 – C1	B1 – C1	B1 – C1
Year of studies	2nd – 5th	5th	2nd – 5th

* Language levels vary between intermediate and pre-advanced and the classification and levels are based on the Common European Framework of Reference for Languages (2004).

Design
The participants were offered questionnaires to check their initial levels of IIC (both dimensions), intercultural knowledge and intercultural skills. The questionnaires were prepared in English and handed out to the students present on the day of testing from all (five) courses of the Faculty of Foreign Languages at TSU.[16] The students were offered the following questionnaires in this order:
- Questionnaire 1 to assess their levels of intra-IIC – adapted from Matsumoto et al. (2001) – see Appendix B-2;
- Questionnaire 2 to assess their levels of inter-IIC – adapted from INCA (2007) – see Appendices B-1 and B-4;
- Questionnaire 3 to assess their levels of critical thinking – adapted from Pintrich et al. (1991) – see Appendix B-5; and
- Questionnaire 4 to assess their levels of self-reflection – adapted from Pintrich et al. (1991) – see Appendix B-5.
- Questionnaire 5 to assess their levels of intrinsic motivation – adapted from Ryan and Deci (1992) – see Appendix B-3.

The questionnaires were paper-pencil based and included different types of questions. The ICAPS questionnaire, and MSLQ questionnaire were based on statements that students could rate on a 7-point Likert scale from 'strongly disagree' (1) to

---

[16] *Only to those that were selected; i.e. who had shown sufficient levels of English as explained above.*

'strongly agree' (7). The INCA questionnaire was designed in the form of open questions which were later coded and graded on a 1-5 point scale.

Before the first testing round in April 2008 (Y1) and in between the other two testing moments in September 2008 (Y2) and March 2009 (Y3), we adapted and improved the questionnaires if needed. The redesign of the questionnaires before starting the first test run in April 2008, is already reported in section 3.5.2 above. We will report in detail on further changes below.

Data Collection Procedures
For all three testing occasions (Y1, Y2 and Y3), the testing sessions took place on different days; each day for one (sub) group of students. Orally the students were given short instructions on how to fill in the questionnaires. The time for working on each questionnaire was strictly limited and took one class of 1,5 hours long. In the curriculum at TSU related to IC, the students have to take two courses: the Introduction to Intercultural Communication, also named Introduction to Specialisation (obligatory for all students of the faculty in the first semester of their study), and Intercultural Communication (also obligatory for all students of the faculty in the ninth semester of their studies). Only the fifth-year students had already taken the second course on IC with a detailed study of the subject, while the other students in the test group were only familiar with some basics of IC from their introductory course. In spite of the fact that the students had different levels in their competency on IC, the basic knowledge available was more than sufficient to fill in the questionnaires, that were not focused on factual knowledge questions in any case.

We tested whether the measurement instruments are reliable (i.e. they test correctly what they are supposed to test) and whether they are internally valid, i.e. that different scales do not test the same construct. The results reported for reliability (or internal consistency) center around Cronbach alpha. Cronbach alpha is the index commonly used for internal consistency of items in a scale. It indicates the extent to which item responses obtained at the same time also correlate highly with each other (Cronbach, 1951). If the value of Cronbach alpha is higher than .60, the instrument is sufficiently reliable for experimental use (Garson, 1998).[17] To indicate whether the measurement instruments are valid, we calculated Pearson correlations. If the Pearson correlation between two instruments is below .80, we consider that they measure a different construct (even though below but close to .80 suggests a partial overlap in instruments).

4.4 Results of testing the measurement instruments

Having constructed the measurement instruments, based on the components and psychological traits favourable to development of ICC, and core aspects of our

---

[17] *The widely-accepted social science cut-off is that CA should be .70 or higher for a set of items to be considered a scale, but some use .75 or .80 while others are as lenient as .60, especially when used for experimental purposes (D. Garson, 1998, 2008). Last updated 9/9/2008. http://faculty.chass.ncsu.edu/garson/PA765/standard.htm.*

model on IIC (see Table 3.1 and Figure 3.1), that have been selected from a wide range of available measurement instruments, in this section we present:

- Whether the selected (and adapted) items are reliable, i.e. internally consistent and testing what they are supposed to test;
- Whether the selected (and adapted) instruments are valid, i.e. that the different instruments do not test the same construct;
- What further changes we have introduced in the different instruments during and following our testing moments (Y1, the first test moment in April 2008; Y2, the second test moment in September 2008 during the trial course; and Y3, the final test moment during the full course run).

Intercultural Adaptation Potential Scale (ICAPS)

At three different moments, we have tested the levels of reliability and internal validity of the items constituting the ICAPS instrument. Table 3.7 shows the reliability results by presenting the different values for Cronbach alpha for the three testing occasions. Table 3.8 shows the internal validity results by presenting Pearson correlations for the subscales of the instrument.

Table 3.7 Reliability indices for ICAPS (values of Cronbach alpha)

	Testing of measurement instruments (Y1) ICAPS-55	*Trial course run (Y2) ICAPS-46*	*Full course run (Y3) ICAPS-46*
Moment 1	.75 (N=95)	.85 (N=31)	.80 (N=90)
Moment 2	.80 (N=95)		.90 (N=94)
Moment 3			.89 (N=95)

With values of Cronbach alpha of .75 and .80, obtained from two measurement occasions during the first trial, reliability of ICAPS was satisfactory. A total of 123 students participated in this first trial, but not all questionnaires were fully filled out, so there are small differences in the number of questionnaires counted for each of the scales of the measurement instrument. For the total measure 104 questionnaires were fully filled in. When looking at the scales within the ICAPS measure, we found that emotional resilience (Cronbach alpha of .74 with 110 observations) and creativity (Cronbach alpha of .60 with 110 observations) were reliable. The scales of flexibility and openness failed to pass the .60 threshold (with Cronbach alpha's of .47 and .17 respectively).

The Pearson correlations, calculated in order to test whether the different scales really measure something else, for the first test show that all scales are correlated at 1% significance level (except for openness and emotional regulation at the 5% significance level). However, correlations between the concepts were not very high,

indicating that the different scales of ICAPS each measure something different, i.e. are internally valid.

Following the first reliability test, before the trial course run, we decided to adapt the ICAPS-55 by deleting nine items from the scale. ICAPS-55 became ICAPS-46. We did this because student feedback clearly indicated that twelve items were not clear to them, leading them to fill in random answers. We studied each of the twelve items and for three items decided to leave them in. For example, "Some people are just no good" (item 17). The other nine items, items 21, 23, 24, 31, 34, 35, 38, 51 and 53, were removed from the instrument, because we believed they were not clearly phrased, students did not understand them, and they did not add new dimensions to the scales. Item 24, for example, was removed: "I am set in my ways", and so was item "People who want to get rid of sexy publications are mentally ill". This meant that the number of items per scale changed as follows: emotional resilience remained with 14 items, flexibility went from 13 to 10 items, openness from 14 to 11 items, and creativity from 14 to 11 items. This makes 46 items in total, but does not change the original Matsumoto et al (2001) scales, just the number of items per scale.

Table 3.8 Pearson correlations between scales within the ICAPS instrument

Testing of measurement instruments (Y1 moment 1) (N=118) ICAPS-55

	Emotional resilience	Flexibility	Creativity
Flexibility	.31**		
Creativity	.32**	.36**	
Openness to experience	.19*	.31**	.07**

Trial course run (Y2) (N=31) ICAPS-46

	Emotional resilience	Flexibility	Creativity
Flexibility	.20		
Creativity	.22*	.34	
Openness to experience	.34	.37	.31

Full course run (Y3 moment 1) (N=97) ICAPS-46

	Emotional resilience	Flexibility	Creativity
Flexibility	.31**		
Creativity	.35**	.36**	
Openness to experience	.41**	.33**	.41**

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level

ICAPS-46 was tested and reliability as measured by Cronbach alpha increased to .85 (Y2), implying that internal consistency of the items – and thus reliability – increased because of the changes. Reliability of the emotional resilience scale stayed the same with a Cronbach alpha of .74, but reliability of creativity (Cronbach alpha of .65) increased and so did reliability of flexibility (Cronbach alpha of .47) and

openness (Cronbach alpha of .43), even though the last two scales are still below the Cronbach alpha = .60 threshold level. The Pearson correlations show insignificant positive correlations between the scales, confirming that the different scales measure different aspects of ICAPS. Following the trial course run no further changes to ICAPS-46 were made. During the full course run, the same measurement instrument was used as during the trial course run. Reliability of the instruments was high, with values of Cronbach alpha of .80, .90 and .89 for the three moments respectively. The highest Pearson correlation between scales of ICAPS-46 was .41 (between openness and emotional resilience and openness and creativity) while all others were lower, from which we conclude that internal validity is sound. Because of lack of reliability of two individual scales, we have decided to use this measurement as a whole and not with the separate scales. As a whole, for measuring intrapersonal IIC, we conclude to have used a reliable measurement instrument.

Intercultural Communicative Competence Assessment Questionnaire (INCA)
Like for ICAPS, we have tested the levels of reliability and internal validity of the items constituting the INCA instrument. Table 3.9 shows the reliability results by presenting the values for Cronbach alpha for the three testing occasions. Table 3.10 shows the Pearson correlations to check for internal validity. During the first testing moment (Y1), we tested for the reliability of INCA and got a level of reliability of the total score, as indicated by Cronbach's alpha, of .73 (see Table 3.9). This is satisfactory, but not too high. Out of the 123 students participating, 110 filled in the INCA questionnaires. We found that reliability of the individual scales was not very high. Only communicative awareness (Cronbach alpha of .62) and tolerance of ambiguity (Cronbach alpha of .60) were above or on the threshold level of Cronbach alpha of .60.

Table 3.9 Reliability indices for INCA (values of Cronbach alpha)

	Testing of measurement instruments(Y1)	Trial course run (Y2)	Full course run (Y3)
Moment 1	.73 (N=100)	.71 (N=29)	.66 (N=94)
Moment 2			.96 (N=85)
Moment 3			.82 (N=83)

Table 3.10. Correlations between factors within the INCA instrument

	Knowledge discovery	*Relatedness to other-ness*	*Empathy*	*Tolerance of ambi-guity*
Y1; moment (N=100)				
Relatedness to otherness	.26**			
Empathy	.23**	.15		
Tolerance of ambiguity	.37**	.36**	.21*	
Communicative awareness	.26**	.36**	.33**	.33**
Full course run (Y3 moment 1) (N=83)				
Relatedness to otherness	.21**			
Empathy	.12**	.18*		
Tolerance of ambiguity	.32**	.39**	.15**	
Communicative awareness	.32**	.27**	.43**	.21**

** Correlation is significant at .01 percent level; * Correlation is significant at .05 percent level.

The other scales were not so reliable (knowledge discovery Cronbach alpha of .47; relatedness to otherness a Cronbach alpha of .37; and empathy a Cronbach alpha of .25).[18] Especially the result on empathy was then investigated further. It turned out that our test panel, the Russian students at TSU, mixed the expression of empathy with traditional Russian hospitality, which led to different answers than envisaged by the instrument designers. Most of the students had the same answers not leaving much variation in the answers. For example, the answer to the open question on the scenario "To invite or not"[19], most students showed their hospitality and not empathy. Or in the other section of the questionnaire, "Business Trip to China", in an an-

---

[18] *We did not measure behavioural flexibility separately as an individual scale, because we only included one item from INCA related to behavioural flexibility in the total measurement. However, this one item was taken into account in the total scale.*

[19] *Scenario 'To Invite or Not': A young person from country X comes to work at your firm for 6 months. You are aware that he is rather isolated and you consider the idea of inviting him to join you and your friends. The problem is that your group of friends have known each other for a long time and a stranger would find it difficult to fit in. Write down in a few lines what you think you would do in this situation (and why). (INCA project, available at: http://www.incaproject.org).*

swer to the question meant to show their empathy, they, again, expressed their thoughtfulness towards the guests instead[20]. Unlike the ICAPS-questionnaire that makes use of a 1 – 7 Likert scale, the INCA questionnaire uses open questions that need to be coded. Inter-coder reliability for the first testing occasion was high with a correlation of .89 between the two coders.[21] Before the first testing moment, we had decided to change the 1 – 3 level of competence scale of INCA to a 1 – 4 scale to create more variance in the answers. This only worked to a limited extent. Pearson correlations of the INCA measurement show that the scales do not measure the same thing within the measurement, i.e. the internal validity of the measurement is demonstrated. We do find significant correlations, but they are not very high.

Following these findings from our first testing of the INCA instrument, we decided to carry through two changes. First of all, we expanded the 1 – 4 coding scale for INCA competences to a 1 – 5 scale. This means we have coded the answers by categorising them as basic – basic/intermediate – intermediate – intermediate/full – and full competence. Second, we adapted the questions regarding the ICC component of empathy to make clear that it is not the same as hospitality, reducing the ambiguity of the questions for the Russian test audience. These changes have resulted in a Cronbach alpha value for the total INCA instrument of .71 during the trial course (29 students filled in the questionnaires). The Cronbach alpha values for the individual scales were as follows: tolerance of ambiguity .62, knowledge discovery .84, communicative awareness .47, empathy .34, and relatedness to otherness .31. This means that reliability of empathy has gone up significantly, but not to the threshold level of .60. Overall reliability has not changed, but some of the scales have become more reliable. The intercoder correlation was .90, showing that both coders of the questionnaires were much aligned in their ways of coding.

Following the trial course run, no further changes to the INCA instrument were made. During the full course run, the same measurement instrument was used as during the trial course run. Reliability of the instruments – measured at three moments during the full course run – was much higher than expected based on the first test and trial run, with values of Cronbach's alpha of .66, .96 and .82 for the three moments respectively. At moment 3, the Cronbach's alpha values for the factors were .66 for tolerance of ambiguity, .49 for communicative awareness, .38 for empathy, .68 for knowledge discovery, and .53 for relatedness to otherness. The highest Pearson correlation between scales of the INCA measurement was .43 (between empathy and critical awareness) while all others were lower, from which we conclude that internal validity is sound. Because of lack of reliability of three (first test) or four (trial run) individual scales, we have decided to use this measurement as a whole and not with the separate scales. As a whole, for measuring interpersonal IIC, we conclude that we used a reliable measurement instrument.

---

[20] *Scenario 'Business Trip to China', question 1: Mr Parker is interrupted in his description of the problem: Wang: Mr Parker, if you should have any problems whatsoever, please let Mr Wang know. We will do our best. What could Mr Wang mean by this? (INCA project, available at: http://www.incaproject.org).*
[21] *One coder coded a random sample of 72 questionnaires; the other coder a random sample of 38 questionnaires.*

Having defined two dimensions of IIC, measured via an adapted INCA question-
naire and an adapted ICAPS questionnaire, we needed to check whether the two
questionnaires were not too strongly correlated, i.e. whether INCA and ICAPS
measure something different, i.e. whether inter-IIC and intra-IIC were not overlap-
ping too much.

Calculating the correlations, we found that the correlation between the two meas-
urement instruments was low and only in some two out of the three measurement
occasions were significant at the 5% level of significance. For the session on testing
the measurement instruments, the correlation between INCA and adapted ICAPS-55
was .23, while for the trial course, the correlation between INCA and ICAPS-46 was
.10. For the full course run, this correlation was .11. This implies that the two in-
struments measure two distinctly different aspects of IIC. These results are presented
in Table 3.11.

Table 3.11. Correlations between INCA and ICAPS instruments

Testing of measurement instruments (Y1 moment 1)

	INCA	ICAPS (55 items)
ICAPS (55 items)	.24	
Trial course run (Y2) (N=25)		
ICAPS (46 items)	.10**	.98**
Full course run (Y3 moment 1) (N=95)		
ICAPS (46 items)	.11	.97**

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level. NA
means 'not available' because in the first test run, we only used ICAPS-55, while in the second and third
testing occasion, we only used ICAPS-46.

MSLQ instrument – critical thinking
Critical thinking, a driver for intrapersonal development as explained in chapter 2, is
a subscale of the MSLQ instrument developed by Pintrich et al. (1991). At three
different measurement moments, we have tested the levels of reliability and internal
validity of the MSLQ-CT. The results are presented in Table 3.12 (values for Cron-
bach alpha).
    The levels of reliability showed a high level of internal consistency as is shown
in Table 3.11 for all three measurement occasions. Before starting the first test of the
MSLQ-CT instrument, we enlarged the number of items from five to seven (see sec-
tion 3.5.2 for the details). This scale enlargement indeed increased the reliability of
the scale compared to the original Pintrich et al. (1991) level of reliability reported
in section 3.3.2.

The Cronbach alpha value during the first testing of the instrument was .82, during the trial course it was .72 and during the three moments of the full course run, the Cronbach alpha was .82, .86 and .84 for moments one to three respectively. The questionnaire thus proved to be reliable in the try-out of the instruments, so it has not been changed afterwards. Given these reliability results, we concluded we had chosen and specified a good instrument to measure critical thinking that suits our research purpose.

Table 3.12 Reliability indices for critical thinking (values of Cronbach alpha)

	Testing of measurement instruments (Y1)	*Trial course run (Y2)*	*Full course run (Y3)*
Moment 1	.82 (N=110)	.72 (N=24)	.81 (N=86)
Moment 2			.88 (N=83)
Moment 3			.85 (N=83)

Self-reflection instrument
We constructed 13 items on self-reflection. As explained in detail in section 3.5.1, part of the self-reflection scale (seven items) comes from the MSLQ instrument, subscale self-regulation – since Pintrich et al. (1991: 23) define part of this scale as "self-testing and questioning" as well as "fine-tuning and continuous adjustment of one's cognitive activities".

Table 3.13 Reliability indices for self-reflection (values of Cronbach alpha)

	Testing of measurement instruments (Y1)	*Trial course run (Y2)*	*Full course run (Y3)*
Moment 1	.81 (N=96)	.74 (N=24)	.87 (N=86)
Moment 2			.90 (N=82)
Moment 3			.87 (N=83)

We added six items ourselves before starting to test the measurement instrument, to increase the representation of the instrument, and the reliability of the scale, keeping in mind the higher order cognitive skill in dimension four of our model of IIC we want to measure. In Table 3.13 the test results for reliability of the measurement on self-reflection are presented. Through all three testing occasions, as the Table shows,

the value of Cronbach alpha was well above .60, which indicates a high degree of internal consistency, with a value of Cronbach alpha of .81 for the testing of measurement instruments occasion, .74 during the trial course and .81, .76 and .81 for the three moments during the full trial course.

Table 3.14 Pearson correlations for self-reflection and critical thinking

	Testing of measurement instruments (Y1)	*Trial course run (Y2)*	*Full course run (Y3)*
Moment 1	.72**	.74**	.66**
Moment 2			.75**
Moment 3			.81**

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level.

Having defined two higher order complex cognitive skills that are expected to facilitate growth of IIC, critical thinking and self-reflection, and having created two measurement instruments for these skills, we needed to check whether the two instruments were not too strongly correlated, i.e. whether MSLQ-CT and the SR instrument measure something different. We expected a correlation to be on the higher side, because both are related complex cognitive skills.

Calculating the correlation, we found that the correlation between the two measurement instruments was significant and rather high as shown in Table 3.14. For the testing session on measurement instruments we found a correlation between the two instruments of .72, while in the trial course, the correlation was .74 and for the final course run, it was .66, .75 and .81 for different moments. This implies that these measurements share variance but still measure distinct concepts.[22]

Intrinsic Motivation instrument
Like for the other variables, at three different moments, we tested the levels of reliability and internal validity of the items constituting the IMI instrument. Table 3.15 shows the reliability results by presenting the different values for Cronbach alpha for the three testing occasions. Table 3.16 shows the Pearson correlations to check for internal validity during Y1, Y2 and Y3.

Looking at Cronbach alpha values in Table 3.15, we see that reliability of the total instrument is high at all moments of measurement, with a lowest value of .75 at the first testing moment.

---

[22] *It is interesting to see that across moments during the full course run, the correlation between the concepts increases as IIC grows. This could mean that when the fourth dimension of intellectual capabilities grows, both skills are more deeply and frequently used, leading to more interaction between the two and thus higher correlations.*

Table 3.15 Reliability indices for IMI (values of Cronbach alpha)

	Testing of measurement instruments (Y1)	Trial course run (Y2)	Full course run (Y3)
Moment 1	.75 (N=95)	.86 (N=31)	.97 (N=95)
Moment 2			.91 (N=95)
Moment 3			.88 (N=95)

Removal of the factor 'pressure' from the instrument following the first test run as described above, increased reliability of the instrument to .86. All factors had Cronbach alpha values of above .60 in this test run, i.e. the changes made to the instrument had improved its internal validity. Without further changes, the adapted IMI questionnaire was used for the full course run, yielding a Cronbach alpha of .88 for the whole instrument and strong values for reliability for the factors. Effort and perceived choice showed the lowest values of Cronbach alpha of .63 and .67 respectively.

Table 3.16. Correlations between factors within the IMI instrument

Testing of measurement instruments (Y3 moment 1) (N=95)

	Related-ness	Inte-rest	Percei-ved com-petence	Effort	Pres-sure	Percei-ved choice	Value
Interest	.57**						
Perceived competence	.03	.34**					
Effort	.51**	.77**	.23*				
Pressure	-.04	-.04	-.29**	.08			
Perceived choice	.49**	.64**	.38**	.58**	-.07		
Value	.49**	.70**	.26**	.67**	.06	.67**	
Motivation	41**	.48**	.17	.40**	.16	.54**	.62**

** Correlation is significant at .01 percent level; * Correlation is significant at .05 percent level.

Turning to internal validity of the instrument, the Pearson correlations for the factors of IMI are presented below. Although some of the factors within the IMI instrument were interrelated and significant to some extent, correlations were not that strong that we should combine factors into one score (see Table 3.16). We therefore could have kept the individual factors separately. However, within our research we were not interested in the details of motivation, but only in the total score. Therefore, based on validity, reliability analysis and the fact that we only need the total score, we decided to keep the factors together for the analysis.

Validity of the measurement instruments
We have already checked for the internal validity of ICAPS and INCA by comparing Pearson correlations, and concluded that their correlation was low, implying the ICAPS and INCA instruments measure different concepts of IIC. We also checked for internal validity of the measurement instruments of MSLQ-CT and the Self-reflection instrument, and concluded they share variance to a large degree, but also that they still measure distinct concepts.

Table 3.17 Pearson correlations on measurement instruments (full course run)

Moment 1

	ICAPS-46	INCA	MSLQ-CT	Self-reflection
INCA	.11			
MSLQ-CT	.20	.25*		
Self-reflection	.26*	.28**	.66**	
IMI	23*	.21	.53**	.57**

Moment 3

	ICAPS-46	INCA	MSLQ-CT	Self-reflection
INCA	-.02			
MSLQ-CT	.40**	.11		
Self-reflection	.42**	.03	.81*	
IMI	.42**	-.09	.56*	.58*

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level.

The internal validity of the intrinsic motivation instrument was also tested and found satisfactory. The correlation matrix in Table 3.17 (figures reported are for the full

course run, moment 1 and moment 3) shows that this is also the case for the other measurement instruments.

We found a small but insignificant positive correlation between INCA and ICAPS-46 instruments at moment 1 and virtually no correlation between the two at moment 3. We also found – as already discussed - a large and significant correlation between the MSLQ-CT and Self-reflection instruments.

We found positive and significant correlations between the Self-reflection instrument, the MSLQ-CT instrument and the ICAPS-46 instrument. This correlation increased from moment 1 to 3, suggesting that – as expected from the theoretical model – there is a positive link between higher order cognitive complex processes in the mind (like critical thinking and self-reflection) and intrapersonal development (intra-IIC as measured by ICAPS). Having tested the correlations between the measurement instruments multiple times, we concluded that their use is validated.

5.   CONCLUSIONS AND DISCUSSION

In this chapter we focused on constructing reliable instruments to measure intra- and inter-IIC and related constructs of critical thinking, self-reflection and intrinsic motivation. The first four variables are relevant to be able to measure growth in IIC along the lines of our model for stimulating IIC of chapter 2, and the last variable matters for course validity and design (chapter 4).

In chapter 2, our model of intra- and inter-IIC resulted in a focus on three components of ICC: critical cultural awareness, cultural stress tolerance and experimentation with the cultural self. These three components have been closely linked in the literature to psychological traits like tolerance of ambiguity, emotional resilience, flexibility, creativity, openness, and cultural empathy. So, in line with our theoretical model, the measurement instruments have to be able to measure intra-IIC and inter-IIC through these clustered components and psychological traits. Because in our model on IIC growth, the higher order cognitive skills of the dimension of intellectual capabilities (dimension four) are at the core of IIC growth, and because – according to Kholodnaya (2002) – a learner has to go through all four dimensions of intellectual development before IIC starts to grow, we decided to measure two of these complex cognitive skills separately. Growth in critical thinking and self-reflection would indicate two things. First of all, that learners are going or have been going through the fourth dimension of our model for intellectual development – having gone through the other three before. Second, that learners are engaging in complex cognitive mental processes as part of their intrapersonal growth.

Many measurement instruments to measure ICC have been developed and used in practical courses on IC over time (Fantini, 2006; Sinicrope, Norris and Watanabe, 2007). Fantini (2006) identifies 87 instruments, but Sinicrope, Norris and Watanabe (2007) find others and our own research has still found many more. Out of these measurement instruments, we have selected and adapted the ICAPS-46, INCA, MSLQ, and IMI instruments to measure intrapersonal development of IIC, interpersonal development of IIC, critical thinking, and intrinsic motivation. We have used a

combination of the MSLQ instrument and items created by ourselves to develop a scale to measure self-reflection, which we name the Self-reflection instrument.

The ICAPS instrument measures the level of emotional regulation, creativity, openness and flexibility of a learner, linking directly to the components of ICC and psychological traits stemming from our model for growth of IIC. The INCA instrument measures the levels of empathy, communicative awareness, tolerance for ambiguity, knowledge discovery and relatedness to otherness, which also links directly to the variables we need to measure coming from our model for growth of IIC. This implies that our measurement instruments are reflective of the theory we have developed in chapter 2, i.e. their theoretical validity is checked and confirmed.

Following the instrument selection, we have tested the five measurement instruments on three different occasions regarding their reliability and internal validity. For each of the measurement instruments we have looked whether to adapt the scales, items or coding ways in order to increase reliability and internal validity. For ICAPS, we reduced the number of items from 55 to 46, based on feedback from the IC participants in the first test run of the instruments. The coding of the open questions of the INCA questionnaire was changed from a 1 – 3 scale as originally envisaged by the project team (INCA, 2007) to a 1 – 5 scale in order to increase variance and allow for a more detailed classification of answers. We increased the number of items of the MSLQ-CT instrument from five to eight in order to strengthen the instrument and increase its reliability and internal validity. We took the original IMI instrument (Ryan and Deci, 1992) and removed five items; three because they were open-ended questions and two because they discussed 'friends' which was not relevant in our context. This resulted in a 40-item instrument. We rephrased the questionnaires in the future tense and changed the word 'activities' for 'course'. In the second test, we removed the factor 'pressure' (five items) as it was not reliable and was defined opposite our own views on pressure, tension and stress, yielding our final instrument of 35 items. We combined part of the MSLQ scale on self-regulation with our own items to create a new instrument to measure self-reflection, the Self-reflection instrument. Finally, we carefully studied the level of English used in the tests, and based on open feedback from the learners participating, changed and adapted semantics in order to increase understanding of the instruments.

Reliability relates to the internal consistency of the measurement instrument. If internally consistent, the resulting Cronbach alphas should be over .60. Internal validity is tested through Pearson correlation analysis to make sure correlations between measurement instruments are not too high; i.e. the instruments measure different constructs. The ICAPS questionnaire (adapted from Matsumoto et al., 2001) was tested for reliability and the value for Cronbach alpha turned out to be between .75 and .90, which is sufficiently high. We decided to use the instrument as a whole, and not also its individual scales. The INCA instrument (2007) showed an overall level of internal consistency of between .69 and .96. Because of the relatively low levels of reliability for some of the scales within the instrument, we decided to use only the aggregate INCA instrument. Since both the INCA and ICAPS instruments are ways to measure IIC, although different dimensions, we tested whether the two instruments were sufficiently independent from each other. Correlations between the instruments show a highest correlation of .24, with other results showing lower cor-

relations, which is sufficiently low to conclude that they measure something distinctly different; i.e. when measuring intra-IIC and inter-IIC, we are measuring something else.

Our adapted MSLQ-CT instrument for critical thinking was tested for reliability and found reliable (Cronbach alpha between .72 and .86). The Self-reflection instrument was also found highly reliable (Cronbach alpha between .74 and .81), indicating a high degree of internal consistency. When running the correlation analysis, we found that critical thinking and self-reflection had a correlation coefficient of between .66 and .75 (with one reading spiking to .81). This is a rather high level of correlation, but still there is a significant part of unexplained variance to warrant treating the two components separately. Reliability of the IMI instrument was high, with .75 in the first test run being the lowest value. Also the factors constituting the instrument were reliable. Pearson correlations showed that internal validity of the IMI measure was also fine.

Overall Pearson correlations for all five instruments showed no significant correlation, except for – the already explained link between – critical thinking and self-reflection. What is interesting to note is the fact that during the first moment of the full course, critical thinking and self-reflection were not significantly correlated with ICAPS (the measure for intra-IIC), but at moment 3 they are, and significantly so. This could be evidence for the fact that students that gained higher levels of intrinsic motivation as the course progressed, also started to engage at deeper levels, showing a higher correlation between the two instruments. Whether we find an interaction effect between intra-IIC and intrinsic motivation is investigated in chapter 5.

Our reliability and internal validity measurements at three different moments in time - having made some changes to the instruments at various points – show that reliability and validity of the measurement instruments have been established. We are therefore confident that ICAPS-46 instrument, the adapted INCA instrument, as well as the MSLQ-CT instrument, the Self-reflection instrument and intrinsic motivation instrument – in their revised forms – can be used to adequately measure inter- and intra-IIC growth as well as growth of critical thinking and self-reflection skills.