Rater Training for the Assessment of Innovation Competencies Development in Higher Education – First Experiences

Helmut Helker, Christiane Stange
Hamburg University of Applied Sciences

Abstract

INCODE – Innovation Competencies Development – is a transnational project addressing the comprehensive goal of finding new curricular concepts for the development of innovation competencies in Higher Education which is part of the European Strategy 2020. At the heart of the project there is a combination of new pedagogical methods with new ways of quality assurance. Innovation competencies are developed in so-called Research Hatcheries which are evaluated by an assessment tool, the Innovation Competencies Barometer. This tool presupposes an intensive training of the assessors to assure the quality of the evaluation.

The paper gives a short description of the INCODE project work and then focuses on the rater training of which first experiences will be presented. Research Hatcheries as part of innovation pedagogy were first developed by the consortium partner Turku – Finland; work on the measurement tool was initiated by Valencia – Spain. Due to the complexity of this instrument an intensive training of the assessors had to be implemented. Hamburg has developed a rater training consisting of Behavioural Observation, Rater Error and Frame of Reference Training. A first application (> 25 participants) took place in June 2013 in Turku. The paper describes some experiences and consequences.

1. Towards Innovation Competencies

Innovation Pedagogy is an approach in Higher Education that - right from the beginning of the study program - tends to develop and unfold those capabilities in students that help them to produce innovative results within their field of study. Innovation Pedagogy tries to bring into accordance the requirements of the labour market for innovative individuals and collectives with the subject-specific learning outcomes in different Higher Education Programs. The development of Innovation Competence therefore may be regarded as a new “cultural approach” in Higher Education (Kairisto-Mertanen, 2012). In general the conceptual meaning of “innovation” varies from “radically new invention” to “mere enhancement of

---

1 Further information on: www.incode-eu.eu
products or processes already existing” and sometimes is enriched by the inclusion of “usefulness” of an innovation to people or organisations. So considering this broad understanding of “innovation” the concept of Innovation Competencies should be defined broadly too. The definition of Innovation Competencies in students therefore has to be multidimensional. In the INCODE project the description of Innovation Competencies comprises an individual, an interpersonal and a networking dimension\(^2\). In INCODE Innovation Competence is regarded as a cluster competence transcending the mere knowledge base of competencies by including skill- and attitude-based attributes that are instantiated in the individual, the interpersonal and the networking-dimension. Constituting components of Innovation Competence within the three dimensions are identified and operationalised by the Innovation Competencies Barometer that is also developed in INCODE.

Innovation Pedagogy tries to develop Innovation Competence using active teaching and learning methods. In the light of the multidimensional structure of Innovation Competence the application of methods that address not only the knowledge-base of students` resources but also the skill- and attitude-based attributes is essential for the development of Innovation Competencies. For this reason the teaching and learning methods used to develop those attributes have to take into account that the learning outcomes reflecting the presence of Innovation Competence in students have to cover those different attributes and dimensions mentioned already.

Considering the complex nature of Innovation Competence and its manifestation in students` learning outcomes an appropriate learning and teaching method has to be applied to develop and unfold all those properties in students. In INCODE a new active learning method was chosen that is considered to be the best for meeting the demands, with a focus on the social aspects: the Research Hatchery (REHA). The operational idea of the concept is to offer a functional learning environment where students, under counselling, can create new information with reliable methods by carrying out research and development assignments from their own university, companies and other organisations. The goal is to combine teaching and learning with research and development activities as well as serving the purposes of working life. So a Research Hatchery is collective learning based on the social constructive learning approach which seems to be most suitable for the development of Innovation Competencies (Kanerva-Lehto 2011).

\(^2\) For the development of the dimensions see Kairisto-Mertanen (2012).
The evaluation of the learning method with respect to the development of Innovation Competence presupposes an instrument that is able to state differences in the development of Innovation Competence in different individuals at different times during the REHA course in a reliable way. The Innovation Competencies Barometer is a tool that was constructed on the base of a broad literature analysis concerning the concept of innovation and was tested and modified during INCODE. It comprises 25 items in three dimensions that are not interchangeable, do not have the same antecedents and correlate just moderately so that they are specified as “formative”, i.e. that the meaning of the concept of Innovation Competence is defined by those items (Marin-Garcia 2013). Due to its complexity Innovation Competence cannot be measured directly as a single variable, therefore it is important to construct and apply a reliable measurement tool – the Innovation Competences Barometer in combination with a special training for application.

To summarise, a main goal of INCODE is to develop an apt tool to measure the expression and the development of Innovation Competencies in students and furthermore to analyse the interrelationship between learning outcomes and learning method by using this tool. Fig. 1 shows these main elements of INCODE.

**Fig. 1: Structure of INCODE**
Innovation Competencies in students as desired learning outcomes of the curriculum are developed in Research Hatcheries; they are assessed by means of the Innovation Competencies Barometer which is complemented by a special training for its users.

In the next paragraphs the structure of this training and first experience with a training session that was held in Turku will be described.

2. INCODE Rater Training

Why do you need a rater training when / before using the INCODE-Barometer? It cannot be taken for granted that raters agree in their ratings when rating according to the same (ICB-) standards. The complexity of the rating task is obvious: attention has to be paid to different modes and foci of observation, on the whole group as well as on the individual group member, on verbal as well as on nonverbal behaviour, on the overt content as well as on the subtext. Furthermore the observations have to be classified and weighted according to the different items in different dimensions on the 1-10-scale of the ICB. In sum, the features of the observation task demand massive parallel information processing by the observer and a systematic method to transfer observations into ratings. Taking into account further individual differences among raters who might differ in experience, language proficiency, self-confidence, commitment to the task, relationship to the ratees, cultural background, etc. without a special training rater consistency, rater agreement and even inter-rater reliability\(^3\) tend to be low. This is shown for instance in Hoyt and Kerns (1999) who found that especially attributes requiring rater inferences – like Innovation Competence does – contain substantial bias variance. Otherwise several studies show that a correctly designed rater training can enhance inter-rater reliability, especially when using combinations of different training methods (Roch et al., 2012; Bernardin and Pence, 1980; Sulsky and Day, 1994; Thornton and Zorich, 1980).

Therefore in INCODE three training approaches are combined: First a Behavioural Observation Training (BOT) to become familiar with the complexity of observation processes. In BOT raters should learn to substitute automatic categorization processes with controlled cognitive processes. Second a Rater-Error-Training (RET) to inform on common rater errors and how to avoid them. Typical errors are for instance similarity-errors (the rater is more favourable to people who are similar to herself/himself), leniency (raters shy away

\(^3\) For a discussion on rater agreement and rater reliability see Graham et.al. (2012).
from negative judgements), halo-errors (raters generalize from one prominent personal characteristic to other aspects of behaviour), central tendency-errors (raters use only the middle range of the scale and avoid extreme judgements), primacy-errors (the first impression has a strong influence on all the following assessments) and context- or contrast-errors (situational factors have a strong impact on the rating results). Finally, a Frame of Reference Training (FOR) to become acquainted with relevant behavioural indicators of innovation competence including examples aligned with the specific context in which the assessment is to take place. In FOR raters discuss their own standards and implicit theories of performance in comparison to normative standards developed by experts. The overall aim is to share a common understanding of performance standards.

Who should participate in the INCODE Rater Training? As the Barometer is developed to be used by teachers and peers for teacher- and peer-assessment and in a third version for self-assessment of the students the target group of the INCODE Rater Training consists of teachers and students.

3. First experiences with the INCODE Rater Training

Participants in the Turku rater training were 12 members of INCODE and 10 Finnish guests. Project members were in various ways engaged in the development of the ICB and familiar with the general concept of this rater training. The guest-participants were lecturers from Turku University of Applied Sciences or other Finnish universities who were in some way or other involved in REHA-projects. They were as yet not acquainted with the measurement tool and had not participated in a formal rater training before.

The five-hour-training-session was conducted in conference facilities that were equipped with all the necessary technique and enough space for group work.

3.1. Description of the training

A general pedagogical principle of the training was to expose participants to relevant situations and let them experience their own behaviour and that of their colleagues first, and then reflect the experiences in group-discussions before theoretical knowledge and scientific terms were added. This principle had been especially pursued in the first part of the training, the Behavioural Observation Training (BOT) and Rater Error Training (RET). Participants were confronted with their individual styles of perception and observation, with the situation of being observed and scrutinized by others, and with their own typical rater errors.
Results of psychological research concerning perception, observation, rater influence on the ratee, and typical rating errors were presented afterwards, and participants got advice on how to cope with rating problems.

In the second part of the training, the Frame of Reference -training (FOR), the general procedure was slightly different. FOR-training was based on short video-vignettes recorded with students from TUAS, KdG, and HAWH. In these videos the students had to accomplish tasks which required different skills connected to Innovation Competence. After a description of the ICB-items the participants first had to find out relevant behavioural clues representative for single ICB-items, then had to rate the behaviour of the students in the following videos on only a few items of the ICB with an increasing degree of difficulty. The last training phase comprised rating the innovation competence of the students in a longer video with the complete ICB.

The procedure was always nearly the same: participants first rated individually, then compared their ratings in small groups, discussed reasons for their ratings, and were then confronted with the ratings of experts. Lastly differences between the ratings were discussed and the experts’ ratings were explained.

The Turku-training ended with a feedback round and a discussion of the training itself. Future trainings should end with a screening test for it is important to determine which persons have really acquired the necessary skills to work as a reliable rater.

3.2. Feedback

All participants gave oral and written feedback. Some points of criticism will be summarized here and considered later in a revised version of the INCODE Rater Training.

Some trainees felt unable to cope with the high number of items especially in the last training sequence when a longer video was to be assessed with the complete ICB. There is obviously a gap to be closed between the mini-tasks of rating only a few items and the application of the complete ICB with 25 items.

Video scenes for assessment training were generally regarded as useful and English subtitles for international groups as very helpful. But videos were also criticised by some for being too short and allowing only for superficial ratings. It will be necessary to communicate that training is a step by step process to build up skills which starts at simple tasks.
Some trainees criticized that there was no video example of networking skills. To find truly representative video scenes for illustrating the skills to be assessed in the ICB is a crucial task and challenge for the construction of the training.

4. Conclusions

At present there are several open questions to answer before the INCODE Rater Training can be standardised. Most of them address the task to deal with potential biases within the framework of the training by implementing apt differentiations. They can be clustered as follows.

4.1. Methodology and Standardisation

Differentiation on individual level:

– Raters who are teachers or counsellors often will have other objectives beside the objective of high rating accuracy (e.g. rating results are considered a reflection of their own teaching skills).

– Some students may be known to raters / teachers already. Therefore the data-base for rating innovation competence will not be the behaviour of these students during project group work alone.

– Teachers / raters doing team teaching communicate about their students. Ratings may be influenced by the opinions of their colleagues.

– Face-to-face feedback situations with the students and teachers may lead to more leniency of the ratings as Levy and Williams (2004) found out for employee evaluation.

– Personal relationships between ratees may influence peer assessment.

– General pedagogical beliefs and attitudes toward measurement tools like the ICB seem to play an important role as Henry et al. (2010) demonstrated.

Differentiation on social level:

– How can cultural differences be taken into account?

– There is only one version of the training program for teacher assessment, peer assessment and self-assessment but the assessments and the consequences of the rating results
obviously differ. A differentiation for these three rating situations which differ substantially is necessary.

− Is it possible to construct more individualised (or modularised) trainings for raters with different pedagogical expertise?

− For familiarising a wider range of people with the REHA-concept and with the ICB, the next step will be to conduct rater trainings on an international level. Therefore general questions would be if it is acceptable to have video examples only in English and if it is necessary to translate not only the ICB but also the rater training manual into different native languages.

4.2. Consequences of rating results

− Results can be used in a summative way e.g. to measure the success of a pedagogical approach like the research hatchery or to gain feedback information about individual development of competences or to change the self-perception of individuals in project groups. Does the further use of the rating results demand a differentiated training structure?

4.3. Further questions

− Is it always necessary to apply the complete ICB? Some participants of the Turku training proposed to use only parts of it depending on the special learning objectives of the actual project or the developmental phase of the project.

− Especially for international projects like INCODE: is an additional monitoring of training sessions necessary as Johnson, Penny and Gordon (2008) proposed? Is it perhaps necessary to videotape rater training-sessions, to compare the ratings of rating sessions in different countries and perhaps train the trainers?

− It is to be expected that in the course of time even very reliable raters will change their rating behaviour. Therefore regular re-training is indispensable. How should these re-training sessions be conducted?

To conclude – more research has to be done to determine the degree of differentiation needed for the training structure. Evaluations of the second round of REHA pilots in INCODE for instance will yield more data that can be used to decide which aspects of the differentiations mentioned are the most important for a modified version of the INCODE training.
References


