



Establishing evidence of construct: A case study

Barry O'Sullivan

Roehampton University

Cathy Taylor

Trinity College London

Dianne Wall

Trinity College London & Lancaster University

Purpose of our paper

- To describe approach taken by Trinity College London to gather empirical evidence of the construct underlying its examinations
- To present main research findings
- To present implications for Trinity and the language testing community

Outline

- Why Trinity decided to carry out this research
- The examination
- The validation model
- The method
- Research findings
- Benefits of going through such an exercise

Purpose of the study

To gather empirical evidence of the construct underlying one of our examinations

purpose for the study

ILTA A test designer must decide on the construct to be measured and state explicitly how that construct is to be operationalised.

Purpose for the study

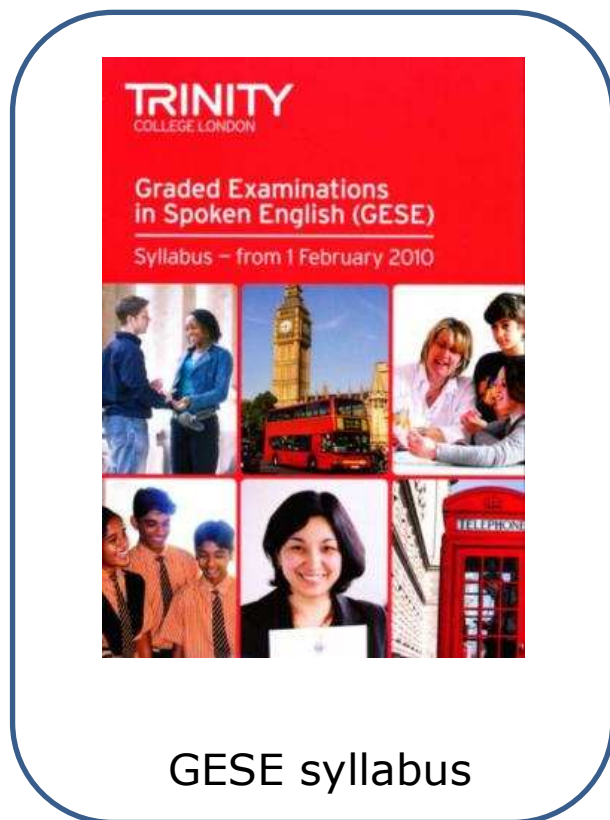
**EALTA Section 1:
Test purpose and specification**

Are the constructs intended to underlie the test/subtest(s) specified?

Section 5: Review

Are validation studies conducted?

Graded Examinations in Spoken English: an outline



Trinity Stages	Trinity grades	CEFR levels
	12	C2
Advanced	11	C1
	10	C1
	9	B2
Intermediate	8	B2
	7	B2
	6	B1
Elementary	5	B1
	4	A2
	3	A2
Initial	2	A1
	1	-

GESE: exam tasks

Advanced 25 mins	Topic presentation	Topic Discussion	Interactive task	Listening task	Conversation
Intermediate 15 mins	Topic discussion	Interactive task	Conversation		
Elementary 10 mins	Topic discussion	Conversation			
Initial 5, 6 or 7 mins	Conversation				

Language skills

- ✓ Communicative skills
- ✓ Language functions
- ✓ Grammar
- ✓ Lexis
- ✓ Phonology

Syllabus requirements

Communicative skills

In the Topic phase

- › Show understanding by responding appropriately to the examiner
- › Communicate a variety of facts, ideas and opinions, and account for these, about a chosen topic linked across a series of extended turns
- › Engage the examiner in discussion of the topic
- › Be prepared to ask and answer questions about the content of the topic
- › Handle interruptions or requests for clarification throughout the discussion of the topic

In the Interactive phase

- › Take control over the interaction
- › Maintain the discourse by asking for information and making comments
- › Help the discussion along by inviting comment from the examiner
- › Take and give up turns when appropriate to do so
- › Where appropriate to the individual task, make use of the language functions listed opposite

In the Conversation phase

- › Show understanding by responding appropriately to the examiner
- › Share the responsibility for the maintenance of the interaction with the examiner
- › In case of a breakdown in communication, show awareness and take basic steps to remedy it

Syllabus requirements

Language requirements

Language functions

- ▶ Giving advice and highlighting advantages and disadvantages
- ▶ Making suggestions
- ▶ Describing past habits
- ▶ Expressing possibility and uncertainty
- ▶ Eliciting further information and expansion of ideas and opinions
- ▶ Expressing agreement and disagreement

Grammar

- ▶ Second conditional
- ▶ Simple passive
- ▶ *Used to*
- ▶ Relative clauses
- ▶ Modals and phrases used to give advice and make suggestions, e.g. *should/ought to, could, you'd better*
- ▶ Modals and phrases used to express possibility and uncertainty, e.g. *may, might, I'm not sure*
- ▶ Discourse connectors, e.g. *because of, due to*

Lexis

- ▶ Vocabulary specific to the topic area
- ▶ Vocabulary specific to the subject areas
- ▶ Appropriate words and expressions to indicate interest and show awareness of the speaker, e.g. *Really? Oh dear! Did you?*
- ▶ Simple fillers to give time for thought, e.g. *well... um...*
- ▶ Phrases and expressions relating to the language functions listed above

Phonology

- ▶ The correct pronunciation of vocabulary specific to the topic and subject areas
- ▶ Rising intonation to indicate interest and surprise as appropriate
- ▶ Falling intonation to indicate the end of a turn
- ▶ Intonation and features of connected speech beyond sentence level

Canale & Swain model of communicative competence

Component	Definition
Grammatical	Knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology
Sociolinguistic	Knowledge of the socio-cultural rules of language and of discourse
Discourse	Ability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances
Strategic	The verbal and non-verbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or due to insufficient competence

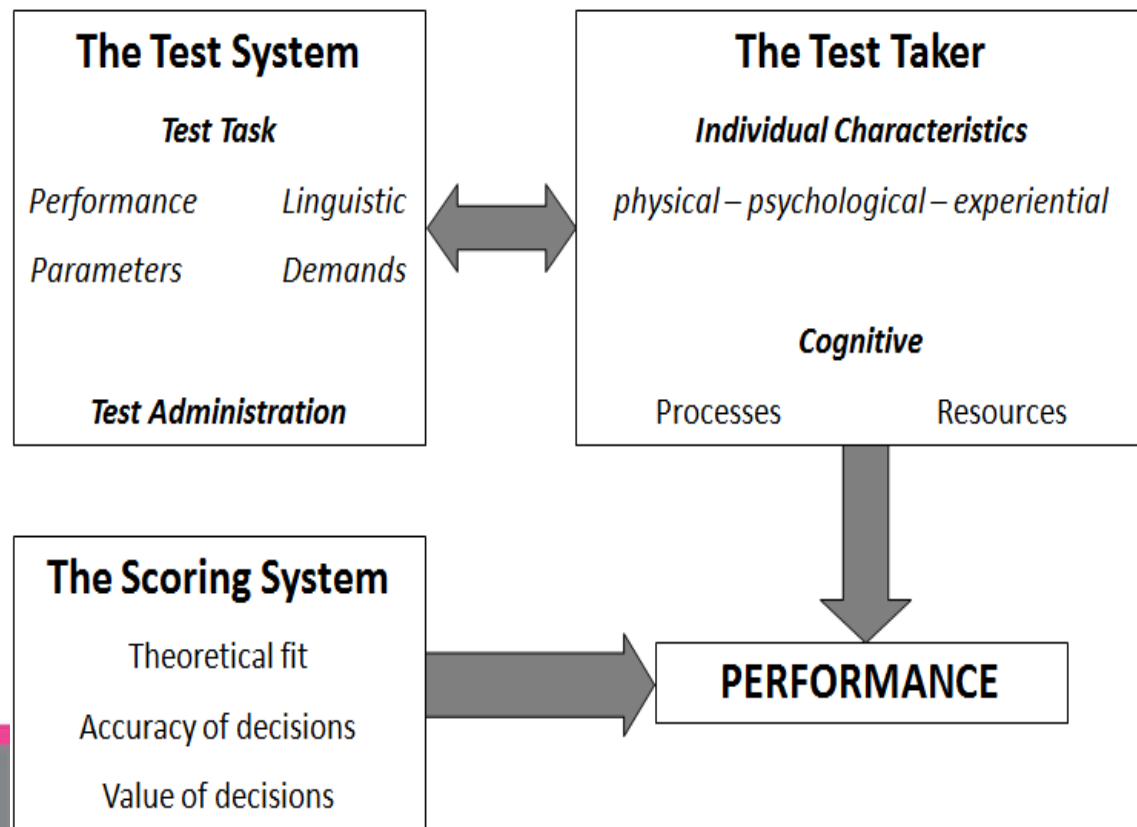
Validity & validation

validity A theoretical model that underpins an argument supporting the use of a test with a specified population for a specified purpose.

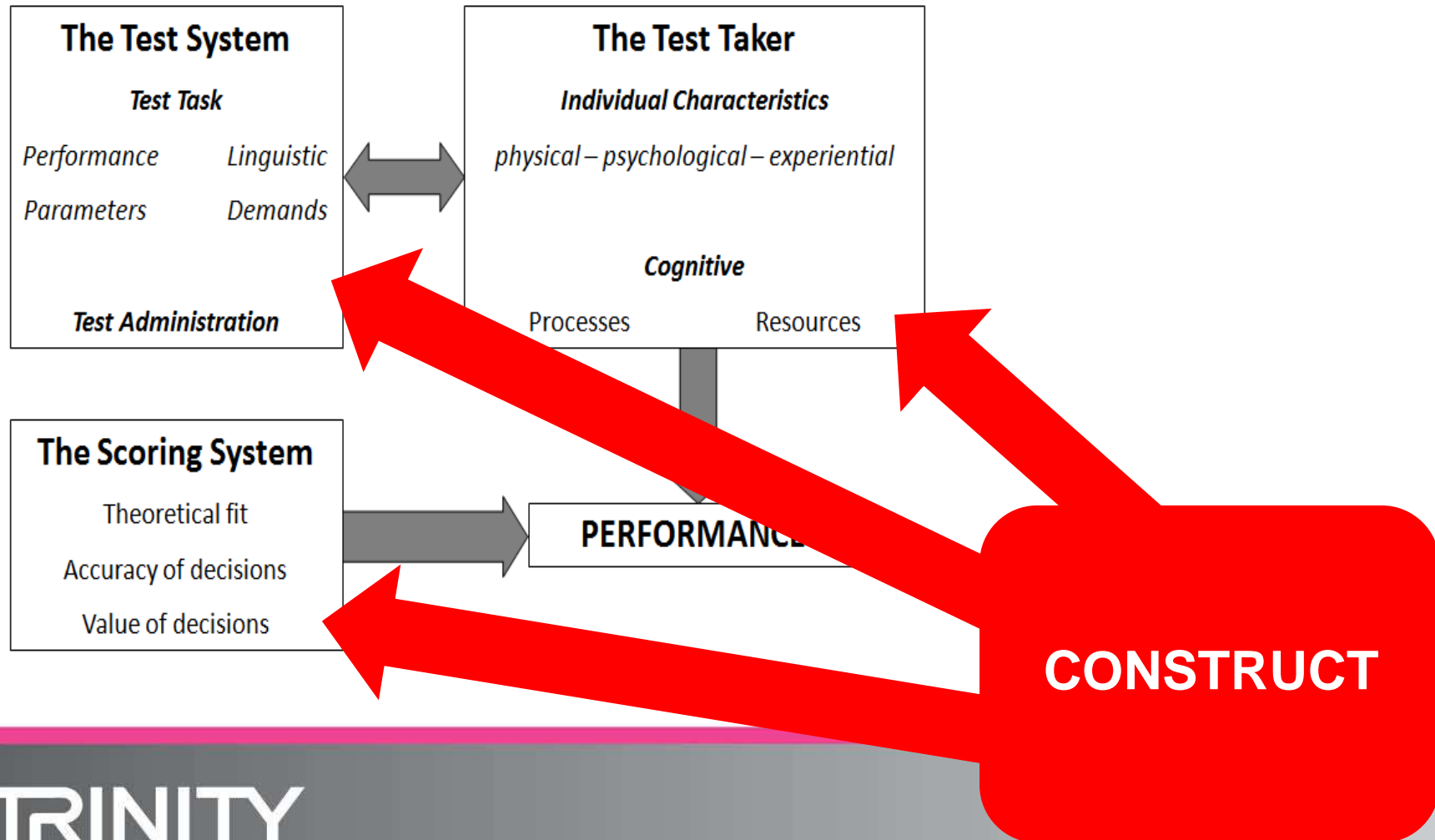
validation The generation of evidence concerning the operationalisation and interpretation of the validity model

The validation model

- First looked to Messick & Mislevy
- Rejected these in favour of the updated Weir (2005) model described by O'Sullivan & Weir (2011) & O'Sullivan (2011)



The validation model



The methodology

1. An overview of the level from the candidate and cognitive domain
2. A specific look at the parameters that define the level (contextual evidence) – this part was divided into a series of tables, one to reflect each of the five task types.
3. An overview of the appropriacy of the scoring system
4. Some general comments on the underlying language model

The test taker I

PHYSICAL	
<i>Short term ailments</i>	Half-fee re-entry possible on application [p.54]
<i>Longer term disabilities</i>	Tasks dyslexia friendly, and the exam can be easily taken by visually impaired candidates. I am not sure what happens if hearing impaired students would like to take the test. The exam sites are accessible for wheel-chair users. The fact that the exam is short and tasks do not last longer than 5 minutes makes the exam accessible for students with ADHD too.
<i>Age</i>	Topic self-selected Not likely to influence performance except perhaps in the interaction phase where certain scenarios might not be appropriate for very young candidates
<i>Sex</i>	Tasks do not seem to show any gender bias

The test taker II

PSYCHOLOGICAL	
<i>Memory</i>	Not applicable [tasks do not rely on memory – though some evidence that memory may be a factor in the topic discussion task]
<i>Personality</i>	Interaction task may put pressure on less outgoing candidates, though there is a measure of interaction in all tasks. The balance with the individual long turn task should make up for this
<i>Cognitive Style</i>	Seems appropriate
<i>Affective Schemata</i>	<p>Dependent on interlocutor. Topics neutral or self-selected and non-problematic.</p> <p>Discussions on the whole are within interest range but even where they are not, the candidates mostly cope well</p>
<i>Concentration</i>	Each section is short enough not to make this a problem – much focus on personal opinion, ideas, experience
<i>Motivation</i>	Most tasks are personalized and motivating and this is an excellent feature of the exam. The only task which might not be as motivating as the others is the interactive task.
<i>Emotional state</i>	Not an issue – topics neutral (though exam nervousness may always be an issue)

The test taker III

EXPERIENTIAL	
<i>Education</i>	Quite a high level of cognitive demand in all tasks so a minimum of secondary/mid-secondary required
<i>Examination Preparedness</i>	Students seem to be well-prepared except for the interactive task, where they do not always understand that their task is to be the initiator
<i>Examination Experience</i>	Does not impact significantly on performance (though see above)
<i>Communication Experience</i>	Candidates I saw appear relaxed with communication needs of test, suggesting that this experience is in place
<i>TL-Country Residence</i>	Not relevant except for certain communication scenarios in the interactive phase e.g. adult college courses

The test taker IV

COGNITIVE PROCESSES	
<i>Conceptualiser</i>	There is on-line planning in the interactive task, pre-task planning in the conversation and topic task, and extensive planning before the exam for the topic task. The tasks vary in cognitive load and difficulty and in the amount of conceptual preparation student do in the exam. This gives a good overview of candidates' speaking ability across planning conditions
<i>Pre verbal message</i>	Student not pressured to respond so likelihood of useful cognitive processing is high Students encode different types of linguistic information in the task and mainly have to rely on procedural and automatic encoding processes. They primarily focus on meaning rather than on form.
<i>Linguistic formulator</i>	
<i>Phonetic plan</i>	
<i>Articulator</i>	
<i>Overt speech</i>	
<i>Audition</i>	Interlocutor speaks at normal rate, no obvious interference from background noise Candidates need to understand a standard British English speaker in the interaction. Speed of speaking and level of vocabulary adjusted to the learners' level but speech still sounds natural. No major misunderstandings occurred in the DVDs I watched.
<i>Speech comprehension</i>	No obvious signs of any issues
<i>Monitoring</i>	Signs of self repair (students mainly monitor content but also some aspects of linguistic form)
<i>Internal</i>	Self-selected topic means that this is well controlled for – interlocutor guides interaction so little value in pre-planning narrative. For other <i>interactive</i> tasks background knowledge may not be relevant
<i>External</i>	In topic-related tasks no significant input except that of the interlocutor – all input provided by candidate. For other tasks little 'new' knowledge presented to candidate. So, likelihood is that external knowledge is not an issue

The test task

TASK TYPE			
Type [title]	TASK SETTINGS		
Format	Purpose	TASK DEMANDS	
	Response format	Input	Interlocutor
	Known criteria	Channel	Candidate Role
	Time constraints	Discourse mode	Variety of accent
	Intended operations	Structural range	Acquaintanceship
		Functional range [use lis	Number of speakers
		Nature of information	Gender
		Content knowledge	
		Output	
		Channel	
		Lexical range	TASK ADMINISTRATION
		Structural range	Physical conditions
		Functional range	Uniformity of administration
			Security

The scoring system

TASK :	
SCORING APPROACH	
<i>Each task scored separately</i>	
<i>Single score for overall performance</i>	
RATERS	
<i>Number of raters</i>	
SCALE	
<i>Type (holistic/analytic)</i>	
<i>Level specific</i>	
<i>Task specific</i>	
<i>Clear distinction between levels</i>	
<i>Based on CEFR/other standards</i>	
<i>Ease of use</i>	

Communicative Competence

TASK NUMBER:	
Strategic Competence	
Discourse Competence	
Sociolinguistic Competence	
Linguistic Competence	
General comments [Reflection on all levels here]	

The findings I

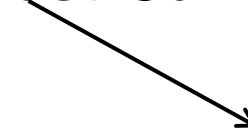
The Test Task



Range of tasks ensure suitability for purpose

Some issues with the listening task

The Test Candidate



Appropriate from all perspectives

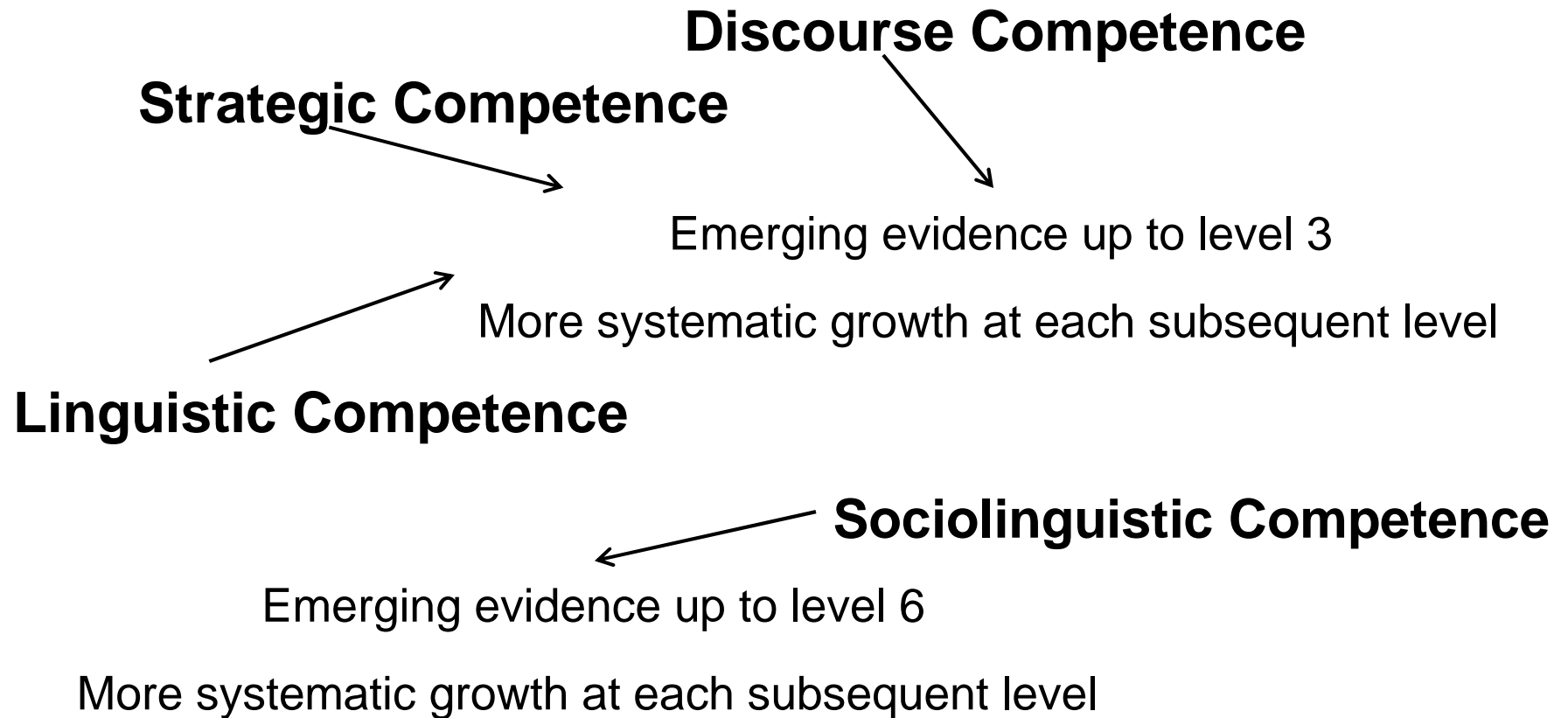
The Scoring System



Procedures appropriate [25% - 30% multiple marking]

Recommendation to reconsider including a specific range of lexis or syntax as rating criteria across all tasks

The findings II



Benefits of the project

- Confirmation of 'communicativeness'
- Confirmed by external experts
- Research projects
- GESE development
- Transferability



Thank you!