

## Chapter 3

### Documentation: Process and Product

Nikolov Marianne and Gróf Szilvia

#### **The Development of the *Detailed Requirements* Document of the Basic-level Examination**

##### **Starting out**

As with all top-down innovative steps in education, the need for the Basic Examination (alapműveltségi vizsga) was triggered by a central decision concerning the development of a new *National Core Curriculum (NCC)* (*Nemzeti alaptanterv* 1995), indicating the typical dichotomy between language planning and language-in-education planning (Kaplan and Baldauf 1997). As a result of the long germination period of over six years, the final – fifth – version of the *NCC* seemed to reflect consensus. As the *NCC* was to be introduced in 1998, the time scale was relatively tight, though the new Basic examination was not to go live before 2002. Despite the fact that the *NCC* includes ten knowledge areas to substitute the traditional school subjects, work on the new exam started along school subjects, reflecting the gap between the wishful thinking of the *NCC* and what is possible to implement in the Hungarian educational context.

In 1996 an official body called OKI AVK (Basic Examination Centre) was established in Szeged to be responsible for the development of the new Basic Examination. This new institution is a satellite of the Budapest-based National Institute of Public Education (OKI) with a schedule and budget dependent on OKI and the Ministry of Education. The aim of the new examination centre in Szeged was to develop documents and the new Basic-level examination for all school subjects identified in the *NCC*. At the same time, new teams at OKI in Budapest started working independently on the two-level School-leaving examination under the label of OKI ÉÉK (for more detail on the background to the examination reform see Fekete, Major and Nikolov, 1999).

##### **Stages in the development of the new Basic exam**

The development of the new examinations involved the following steps:

- First, general requirements were to be developed and after sending them to a variety of professional bodies and schools, to be accepted. This was accomplished for all subjects involved in the *NCC* in the spring of 1997.
- Then, based on the general requirements, detailed requirements were to be compiled, again to go through professional bodies. The process was to finish in 1998, or 1999 at the latest.
- Finally, based on the *General Requirements* and the *Detailed Requirements* documents new tasks were to be designed, piloted and put in a bank. To be able to do so for English a *Guidelines for Item Writers* was developed, while other subject areas did not involve this step, but based item writing simply on the *Detailed Requirements* or used a variety of tasks from previous projects.
- The first piloting of the Basic exam tasks was implemented in the spring of 1999.

As both the Basic examination and the School-leaving examination were to be developed to a similar time scale and within the frame of OKI, it seemed reasonable for people responsible for the projects to co-ordinate the work jointly. As a first step, the first version

of the general requirements of the Basic and School-leaving examinations was commented on by experts of both teams working on the two documents.

In the May 1997 special issue of the journal *Új Pedagógiai Szemle* the *General Requirements* in all subjects of the Basic examination were published with a critical analysis of the feedback from teachers (Nikolov 1997). Later work on the *Detailed Requirements* was based on this source, though a slightly different version of the *General Requirements* was published in the *Művelődési Közlöny* (1997), at that time not known to any team member. This version included less detail, was shorter than the other version and left more questions open.

As the task of compiling detailed requirements required specialists, two teams were recruited in 1997: one for the Basic examination and another one for the School-leaving examination. These two teams collaborated on the tasks for a while, mostly with the support of an external body, the British Council, Hungary.

Work on the *Detailed Requirements* started by the following activities. In 1997 all documents available were carefully studied. They included

- *The National Core Curriculum*
- School-leaving examination documents and test booklets of previous years
- European documents, among them *Modern Languages: Learning, Teaching, Assessment. A Common European Framework of reference* (1996), *Waystage* 1990 (van Ek and Trim 1991), *Threshold* 1990 (van Ek and Trim 1991)
- Draft versions of detailed requirements of the Baltic States
- The published documents of the Slovenian school-leaving examination
- Published booklets of international examinations, e.g.: UCLES, TOEFL
- Documents of the Hungarian state language examination
- Published course materials widely used in Hungarian state education.

After careful analyses of the available background materials and documents, we aimed to compile a document containing the following parts:

- An introduction to the *Detailed Requirements*, including the *General Requirements*, with a rationale, aims and objectives, the structure of the exam, timing and weighting of components, and administration details
- Detailed requirements for the four basic language skills, reflecting the minimal and Basic levels identified in the *NCC*
- Text types and task types for all four skills
- A variety of piloted sample tasks for each skill, hopefully for each task type listed under the four basic skills.

The *Detailed Requirements* document developed through various versions:

- The October 1998 version of the *Detailed Requirements* included text and task types according to the four skills, as well as an introductory chapter.
- The Spring 1999 version included unpiloted sample tasks, and also 'can do' statements on what testees were supposed to be able to do on the minimal and Basic level. This distinction had to be made because the *NCC* prescribed requirements on these two, hardly distinguishable levels. It also included a controversial table on levels of the Basic and School-leaving examination compared to state language exams and European levels.
- In the final, Autumn 1999 version (Cseresznyés et al. forthcoming) the distinction between minimal and Basic level was kept but both the table and the sample tasks were excluded.

- The final version had been commented on by experts in other modern languages as well as English, altogether by five specialists, whose suggestions were integrated into the final document.

It was not possible to omit the above mentioned distinction between minimal and Basic requirements, although work on so-called 'frame curricula' started at the Ministry of Education in September 1999, and as part of this process the 1995 *NCC* is also being revised. As a result of this revision, most probably this problematic distinction will be dropped, so there will not be a need for the *Detailed Requirements* to include two levels for the Basic Examination any more. Despite this recent development, as officially the revision of the *NCC* has not been finalised by January 2000, the final version includes requirements on two levels.

### **Problems with the *Detailed Requirements***

During the years of working on the *Detailed Requirements* and the new exam several problems emerged for a variety of reasons:

- Ownership has been problematic throughout the process: team members felt they belonged to a bigger team of experts working on not just a document, but also the Hungarian Examination Reform, involving all three levels: Basic, Intermediate and Advanced, the first being officially the exclusive territory of OKI AVK (Szeged), whereas the second and third were that of OKI ÉÉK (Budapest).
- English was the only subject in which collaboration across the two OKI institutions was initiated and implemented, but this bottom-up initiative never achieved official acceptance, even though the official agreement with the British Council was signed by BOTH Examination Centres. Other subject area experts found it unnecessary to consult colleagues of other teams, and there was no internal structure or perceived need for official co-operation. The status of experts involved is necessary to elaborate on here: at OKI AVK no subject expert has been hired on a full time basis, but all have been working on contracts, the director included, with full time jobs elsewhere in education. On the other hand, all subject team leaders responsible for the School-leaving exam have had full time jobs at OKI ÉÉK, thus reflecting the difference in the prestige of the positions. It is to be added, however, that full time people have been more threatened for the last two years because of imbalances in educational policy. Educational experts in such posts would lose more than others working on contracts.
- The situation has been further coloured by the fact that the team for the Basic exam was to develop the *Detailed Requirements* for all modern languages, whereas item writing was to proceed in English only with the support of the British Council, as a separate team was working already on German as a foreign language. The underlying official assumption in the earlier stages of the Basic Exam was that English would serve as a model for other languages, and the process should be kept on a low budget. To counterbalance the lack of other modern language experts on the team, external readers of the *Detailed Requirements* document involved specialists of French, English and Italian to ensure legislation and acceptance. With the position and role of the Basic examination becoming unclear under the new government, it was suspected that perhaps the investment in other languages would be a waste, as recently, even the need for a Basic exam has been questioned by the extension of compulsory education from age 16 to 18.
- The Hungarian team members working on the documentation and the development of the new Basic exam were initially inexperienced and enthusiastic, but they received considerable training through the British Council, both in Hungary and overseas. The team expanded from the original six to eight; then the Basic exam team became part of a bigger group of about 15, as a result of the joint Year 10 and Year 12 Project. Later, new item writers were carefully selected and trained by the British Council to

write items for both Year 10 and Year 12 exams, such that, in the summer of 1999, item writers outnumbered original team members.

- At the beginning of the writing of the *Detailed Requirements* a smorgasbord of materials and resources was available, but it was extremely hard to identify criteria along which to make choices. Also, without piloted tasks it was impossible to justify claims and choices.
- A general lack of empirical evidence characterised the process: there was no reality apart from individuals' own experience to use as anchors, but most of the team members had teaching experience in contexts either way beyond the supposedly typical 'Basic level' classroom or no teaching experience at all in state education.
- Double ownership has characterised the process: as team members, we were responsible for the joint document we worked on together, however this was first ignored, and then later actively disliked by both OKI headquarters, and it was unheard-of in other subjects. OKI AVK suggested and implemented changes which the Project as a whole profoundly disagreed with and found unprofessional. These unacceptable changes, which OKI AVK did not discuss with Project leaders, were supposedly intended to bring the *Detailed Requirements* document for modern languages more in harmony with the documents for other subjects.
- One crucial issue has always caused concern throughout the process. External assessment has been perceived as crucial for the English exam, while in other subjects this was seen either as not necessary at all, since objective testing was ensured in mathematics and sciences, or an unnecessary complication and an extra cost, because some people interpreted it as indicating a lack of trust in teachers.
- Also, the time necessary for the development of the *Detailed Requirements* documents for other subjects was much less for two reasons: there was more accumulated experience behind them and they relied on previous documents heavily without introducing innovative ideas.
- Incentives for the writing of the *Detailed Requirements* were low. In other subjects typically one person buckled down to work and came up with a booklet within a few months, while the document for modern languages went through a lot of changes even between official versions, altogether eight team members contributed to it, and much more time and effort was necessary for achieving consensus.
- Finally, the role of the foreign agent needs to be mentioned. At the beginning of the Examination Reform Project unanimous enthusiasm characterised the cooperation between team members, OKI headquarters and the British Council representatives and British advisors. During the implementation of the project a variety of problems have surfaced. Among them the most important ones need to be mentioned:
  - As the posts of some decision makers became gradually unstable due to political changes, they could not, or did not want to risk to, give enough support to the project. To illustrate the point, what started out as an Examination Reform Project contract between the Ministry of Education and the British Council, was finally after over a year of hesitation signed by OKI officials and narrowed down to item writing. These facts have definitely lowered the status of the original project plan.
  - Some team members voluntarily dropped out of the project, because of disagreement over financial matters and serious management problems at OKI.
  - Management has been constantly problematic: changes on the political level kept interfering and the original schedule has been extremely hard to stick to. To illustrate with an example, the political decision to increase the school-leaving age has resulted in a professionally highly problematic pilot project (see Chapter 15 on piloting Basic exam tasks), where tasks intended for Year 10 students were trialed on mostly 8<sup>th</sup> graders.

- Suspicion and jealousy have also emerged as some stakeholders perceived support as interference, and English has gradually been perceived as threatening the status of other modern languages, in particular German. This shift in perception has resulted partly from the lack of support from foreign agencies to other languages, and partly because of the accumulation of expertise and professionalism in the team working on English.

## **Conclusion**

It would be hard to predict to what extent the final version of the *Detailed Requirements for the Basic Exam* (Cseresznyés et al. forthcoming) can be maintained in the long run in the light of future pilots. At the time of writing and rewriting them not enough piloted tasks designed according to the document were available, as there was a lack of time and not enough tasks were available to pilot. Therefore the final loop in design is still missing: the adjustment of the *Detailed Requirements* according to empirical findings. But see Chapter 15 for the empirical results.

Finally, it is time to take stock and summarize what role the *Detailed Requirements* document has fulfilled in the development of the new Basic exam, and how we have benefitted from the process described in this chapter. First, we have developed an expertise which can be applied in further projects in Hungarian education, for example the design of frame curricula, in writing tasks and helping item writers and teachers, and in training markers. Second, a document is now available that can be used as a reference point and is there for further versions to build on according to decisions made and steps taken in educational policy. Third, the principles and criteria, requirements, text and task types are most probably realistic, therefore useful for further piloting. Fourth, as the next stage it will be possible to find empirical evidence to what extent the levels described in the *Detailed Requirements for the Basic Exam* are realistic for the target population of Hungarian students, whether the task and text types match the expected levels, and how students, teachers, parents, and other stakeholders react to such a document. Fifth, in a few years' time it will be possible to look back to see how many of the ideas for external assessment and teacher development will have turned out to be wishful thinking or realistic dreams.

## **Revising the *Guidelines for Item Writers***

### **The reasons for the need to revise the *Guidelines***

The *Guidelines for Item Writers* (see Appendix III) is, to our knowledge, an unprecedented and to this day unmatched document and achievement in Hungary, and is of great help to item writers as well as to those evaluating items. Nevertheless, by the summer of 1999 it became apparent that Version 3 of the *Guidelines for Item Writers*, dated October 23, 1998, needed thorough revision, for four main, overlapping reasons:

1. The new version of the *Guidelines* had to reflect the experience gained from writing items, dealing with items submitted and giving feedback on them as well as our newly-gained knowledge of what makes items good, and from the results of the piloting.
2. It had to be in agreement with our current concept of the School-leaving examination and the documents being revised (indeed in some cases it did not even conform to Version 7 of the *Specifications*, dated the same month).
3. It had to be made more user-friendly.
4. Version 3 contained several inconsistencies, ambiguities, inadequacies, irrelevancies and contradictions, which was probably due to the fact that it was the product of several people's work. Obviously these had to be eliminated.

### **The sources for a revision of the *Guidelines***

- In 1999 the British Council organised training for new item writers on two occasions: a weekend session in spring, and a one-week training course in July. At these sessions participants analysed and evaluated tasks submitted by item writers, the piloted tasks of the Examination Reform Project as well as tasks from already existing public examinations in English such as the Cambridge examinations. We also gained insights into how and on what principles UCLES worked. All this provided invaluable knowledge for us for writing items as well as for giving guidance to item writers.
- As most of the OKI Year 12 team themselves wrote items, together with the other item writers, we had hands-on experience of what it was like to work according to the *Guidelines*, what were its weak points, which of its requirements were feasible and which ones were not, etc.
- In June 1999 item writers submitted about 200 tasks testing all the skills at all levels, and the OKI Budapest team gave feedback on these items. This task revealed many of the weak points in the *Guidelines*, and drew our attention to the need to give more guidance to item writers.
- We also learned lessons from piloting tasks. One such lesson, for example, was that students' anxiety in exam situations might cause them not to do what they have to, therefore the rubrics and the examples must be clearer and articulate better what the task is, and everything should be as standardised as possible.
- Our knowledge of the teaching and testing of different skills as well as further reading of related literature also fed into the work of revising the *Guidelines*.
- The changes made to the *Detailed requirements (Az érettségi vizsga részletes követelményei)* document and the likely changes to be made to the *Specifications (Vizsgaleírás)* document (finished in February 2000) also led to several changes in the *Guidelines*.

### **The process of revising the *Guidelines***

The work on revising the *Guidelines*, which also involved translating it into Hungarian for OKI Budapest, lasted for four months. The first comments on some of the requirements were written down when we were trying to design items to be submitted in June 1999. Then, after giving feedback on the 200 tasks handed in by item writers, the first list of suggestions for changes was drawn up. This was discussed and expanded by the Year 12 team. On the basis of this discussion, the first draft of Version 4 was prepared. This was read and commented on by the Year 12 team as well as by Charles Alderson. Item writers were also asked to send in their comments and suggestions, which, probably due to their busy time at school, they were unable to do and thus we were not able to use their input. In the process of translating the *Guidelines* into Hungarian for OKI, which obviously involved a very careful reading of the document, further questions came up, which were duly discussed by the team. During the autumn of 1999 the *Részletes vizsgakövetelmények (Detailed requirements)* document was revised by the Year 12 team, after which each chapter of the *Guidelines* was checked against the appropriate chapter of the former. During all this work one person was responsible for keeping a record of the outcomes of our discussions, writing up the changes and then re-writing the whole *Guidelines*. In this way it was ensured that the new version was more consistent and coherent.

### **The changes made to the *Guidelines***

It is impossible to list on these pages all the changes made to the *Guidelines* but a detailed account of the reasons for change, with some examples of the problems and their solutions, is given below. Please note that no attempt has been made to put them in any order of importance.

- *Some of the requirements in the Guidelines turned out not to be feasible.*

Version 3 asked item writers to trial items with one or two colleagues and a group of ten students and submit their results along with the task. As it happened, few item writers were lucky enough to have the appropriate group for trialling items (and also to have adequate photocopying facilities), and even if they were, the Editing Committee could not make any use of the results obtained in this way. We realised that it was the item writer who could profit and should profit from this small-scale pre-testing, and a lot can be learned from pre-testing items with colleagues only. Thus Version 4 asks item writers to pre-test their items with two or three colleagues and, (only) if possible, with a group of 10 students, and to then revise their items themselves on the basis of the results obtained. A technical problem arose from requiring item writers to submit original texts rather than photocopies. Items submitted with original texts (e.g. brochures, newspapers, etc.) were difficult to handle. Besides, item writers often did not keep a copy for their own later use, which caused a problem when they wanted to re-submit their revised items. Thus, Version 4 asks item writers to attach either the original text or a good quality photocopy of it, and reminds them to keep a copy for themselves.

Version 3 seemed to require too many sentences for lead-in and lead-out in gap-filling tasks (two sentences at the beginning, preceding the sentence containing the example, and two at the end). This sometimes resulted in quite long chunks of text without items, a low number of items per task, or the rejection of otherwise good texts/tasks. For Use of English tasks, where the text must not be longer than 300 words, this requirement made text selection extremely difficult. We agreed that, as a rule, one sentence was sufficient for lead-in and lead-out.

- *Difficulty of dealing with a large number of items submitted at the same time*

The number of tasks submitted in June 1999 was around 200, the number of revised items submitted in September was around 160. The Editing Committee had great difficulties when trying to keep account of them, preparing and distributing copies of them among the people evaluating the tasks. It turned out that more identification was needed for tasks, therefore item writers are now required to write their numbers and also a task identification on every single page.

- *Item writers not following the Guidelines – Inconsistencies in the Guidelines*

The difficulty of dealing with the large number of submitted items was increased by the fact that many item writers had not followed the *Guidelines*. A large number of tasks was handed in without standard cover sheets, with pages not numbered and labelled properly, without copies of the original texts, without bibliographical information as to the source of the text, without tapes or tapescripts for Listening tasks, without model answers for Writing tasks, etc. In one or two cases item writers submitted several tasks with only one cover sheet for all. Similarly, in many cases task-specific requirements were not adhered to.

Although all this was partly due to the item writers not reading the *Guidelines* carefully enough, we realised that we had to look for ways in which to give more guidance to item writers as to what was required of them. A thorough re-reading of the *Guidelines* revealed that it was inconsistent in several places, e.g. it mentioned certain requirements in one or more chapters but not everywhere where it would have been appropriate. We agreed that each requirement should be repeated in each chapter where it was appropriate even if they were listed in Chapter 2 (General guidelines), thus giving more emphasis to technicalities and other requirements. We also compared the guidelines for those task types which could be used for assessing different skills (e.g. gap-filling for testing reading and use of English) and similarly the guidelines for those skills the testing of which may have some common features (e.g. listening and reading, reading and use of English, speaking and writing) and made sure that all the common requirements were given everywhere as appropriate.

Another reason which might have accounted for item writers not following the *Guidelines* was that the meaning of the terms in the first three headings (1. Task design, 2. Text selection and 3. Item types) was unclear, and thus some guidelines sometimes appeared

under different headings in different chapters. All this made orientation in the *Guidelines* difficult for item writers and indeed for us as well. The new headings we agreed on were 1. Text/Prompt selection, 2. Task selection, and 3. Task design, 'task selection' referring to the selection of task type only. All the requirements belonging here were then re-arranged under these headings.

- *Contradictions in the Guidelines*

Some of the guidelines given were contradictory. The most obvious one had to do with what changes item writers were allowed to make to texts. Whereas in one place item writers were told to indicate whether the text was adapted or edited, in other places they were told not to use adapted texts and not to omit lines or paragraphs. Our experience has shown that in most cases when item writers *had* made some changes, it either made understanding the text more difficult because e.g. the text had become less coherent or it resulted in incorrect sentences. We felt that in our setting it was safer to ask item writers not to make *any* changes to the original text but to give suggestions for changes separately as comments.

- *Inadequacies in the Guidelines*

In Version 3, lists of reference books were not given for some of the skills and where given they had to be rewritten or complemented. Also, some of the tasks submitted were found inadequate for reasons that were not stated in the Guidelines. Thus, for example, some tasks were rejected because they had too few items (Version 3 did not state the minimum number of items that a task must have). We agreed that there should be a minimum of five items in a task (10 in Use of English tasks), otherwise the task would not be economical to use.

In some places the *Guidelines* gave insufficient instructions as to what was necessary for an adequate task, especially in the case of the testing of writing and speaking. For example, Version 4 now asks item writers to avoid Writing and Speaking tasks which require creativity and imagination from students and to provide appropriate and sufficient prompts instead. For the Speaking Paper they are asked to design role-plays, simulations and problem-solving tasks which involve real interaction and have an outcome, and to avoid tasks where the interlocutor has to speak more than the candidate or where the two candidates talking to each other have unequal roles. Similarly, it was necessary to add some more guidelines for some of the task types. For example, gap-filling items in Listening tasks must be based on factual information, not on subjective judgement. Or, when selecting an extract for a Reading task, item writers must make sure that it does stand alone. In intrusive word tasks for testing use of English, there must be only one item in a line, and the word to be deleted must not appear in the line more than once, and must not be the first or the last word in that line. These are all guidelines that were missing from Version 3.

The fact that students often misunderstood tasks in the piloted papers made us realise that more and clearer guidelines were needed as to how item writers were to mark items and options, where and how to give examples, how to lay out tasks, how to formulate and what to include in rubrics, etc. Also, candidates must be given all reasonable help to complete a task, and thus e.g. item writers are now instructed to begin texts with their titles or headlines, and inform candidates in the rubrics for Listening, Reading and text-based Use of English tasks of the text-type and/or source (preferably both). For the same reason, the Guidelines state that the language level of rubrics, of texts for Use of English tasks and of some types of items as well as that of input texts for Writing and Speaking tasks must be below the tested level. Item writers are also asked to organise the requirements in Writing tasks in bullet points so that candidates can follow them more easily.

- *Ambiguities and unclear sentences in the Guidelines*

When we were translating the *Guidelines* into Hungarian, some of the sentences turned out to be ambiguous or unclear, i.e. the members of the Year 12 team interpreted them



differently. These sentences needed to be either rewritten or left out. Examples include the requirement for Use of English tasks: 'Test a variety of items in one task.', which was re-written as 'Test a range of different linguistic points in one task', or the sentence 'Make sure there is only one way of giving answers', which was left out as we interpreted it in different ways, and its possible meanings were expressed elsewhere.

- *Irrelevancies in the Guidelines*

Some guidelines in Version 3 were not relevant to the paper in question, and this might even be misleading for item writers. For example, the chapter on writing contained guidelines for designing gap-filling tasks, which will not be used for assessing writing skills, or the chapter on listening asked item writers to supply unambiguous illustrations. Such guidelines had to be deleted.

Other guidelines became irrelevant because of the changes made to the *Detailed requirements*. Thus, for example, it was decided not to include form-filling tasks in the Writing paper and mediation in the Speaking paper, therefore the guidelines referring to them had to be deleted.

### **Further work envisaged**

The *Guidelines for Item Writers* is by no means a finished document, and perhaps it never should be, at least not until work on the Examination Reform Project has come to an end. It will have to be constantly revised on the basis of the decisions made on the future school-leaving examination in English as well as on the basis of what we learn from evaluating submitted and other tasks, piloting items, item writer training sessions, item writers' and other professionals' feedback, etc.

In addition, at a later stage, the *Guidelines* could be made even more user-friendly by supplementing it with tables giving an outline of each paper of the examination, and by gathering task-specific requirements under separate headings. A glossary of terms might be of great help to new and/or less well-trained item writers. A list of non-desirable topics, however difficult it might be to draw up, would provide useful guidance in text selection. Standardised rubrics will also have to be written and incorporated.