Impact evaluation

Antoni Castelló
Department of Basic, Developmental and Educational Psychology
Autonomous University Barcelona

---

1. Introduction

When an instructional process takes place, there are normally two different perspectives in the analysis of its contribution. On the one hand, the sociological approaches assume that education works and the goals that are declared are also achieved by those who overcome the exams. Instructional effects are seldom evaluated out of the Educational System, and quality concerns normally refer to the correspondence of the educational goals with the instructed materials or the marks obtained by the trainees.

On the other hand there is a, by far, less widespread approach that focuses on the actual learning achieved by the individuals that received the instruction. It does not mean an assumption that education does not work, yet it differentiates educational actions from their effects on learners. The stressed component is learning, that is to say the changes that have taken place in the learners' brains. For this reason it is usually considered as more psychological and, since it focuses on learning, evaluation takes place both within and outside educational settings, stressing the non-academical contexts. I also infers educational quality through the changes in the trainees knowledge structures and behaviour that last far beyond instruction has ended. *Transference* of knowledge and, more recently, *competence* gain, are terms associated to this perspective.

The so-called psychological approach has generated some degree of transformation in the educational-sociological perspective. The ideas, nowadays common in education, that refer to generalisation of the instructed materials and the focus on behavioural competencies, rather than on mere knowledge retrieval, have been theoretically grounded on the psychological approach. However, most of the claims that have boosted the changes in the sociological approach came from the needs that have been made explicit at professional and industrial settings: people entitled by the educational system that were not proficient in their workplace.

That made impact evaluation, which is a natural procedure from the learning-oriented approach, being adopted by the sociological approach as an external validity source. Although these are good news, external validity sources are a bit complicated to establish. What should be the capitalization of instruction? Typically, instruction is also associated to certificates and grades. Hence, if the grade or certificate is a necessary condition to access a set of workplaces it cannot be said neither that the instruction received, nor the learning achieved, is what makes the difference. Consequently, in cases like these, the actual competence and the grade might not match, at least in some individuals. The key point is whether the impact evaluation pays attention to individual behaviour (thus, focus upon the very competence) or it rather infers the competence from other indexes, like the grades.

When competence is inferred there are plenty of situations that may distort the truthfulness of the inference. A few illustrations can shed some light. As a first example, let's suppose that a given professional level shows a higher proportion of men (or women, it does not matter the gender). Concluding that men (or women) are more competent for that job is usually a mistake, because gender-related factors, which have nothing to do with competence, may also be involved. A second example refers to professions where the access is mediated by a tough examination process. In the meritocratic perspective those that pass the exam are the competent ones. And that is true, at least in the sense that they are competent for overcoming the exam, but not necessarily for professional performance. It is going to depend on the distance between the examination procedure and the skills and abilities that support the exercise of the profession. And a final example: when competence is defined according to the

behaviour of the majority of the professionals. It is quite obvious that competence has to do with the appropriate fulfilment of a task. A majority of people not fulfilling a task properly do not convert their behaviour into competent; it would be better described as majority incompetence.

Solid indicators of impact involve behaviour directly. They do not infer it. Inference should be done when the inference procedure is cheaper than the access to real behaviour, and the inference procedure has been clearly demonstrated to work by comparing its results with a solid behavioural approach.

The present document deals with these topics, framing the underlying processes and explaining how they operate. It makes the general reasoning specific for the Serbian Judicial Academy, concentrating on the evaluation and improvement of the instruction taught.

2. Feeling of knowing and changes in behaviour

The effects of any instructional action have different layers of incidence in the instructed persons. Indeed, what a person feels can be an erratic indicator of the actual learning he or she has gained. The learning process itself takes time and it is far from being a direct consequence of the instruction process. When learning consists in complex knowledge structures or even in complex reasoning and behavioural products —usually grounded on knowledge structures— many actions have to be performed by the learner, on his or her own. The typical process consist of the connection of new materials with existing knowledge (both declarative and experiential) as well as a thorough inspection and testing of the new knowledge structure. If behaviour is involved, as it usually happens when competencies are set up, it is a common situation to need to inhibit former response mechanisms and gradually replace them with new ones.

The feeling of understanding something (or even mastering it) is something that depends on the overall knowledge available, the internal coherence of the new materials and the specific framework established in the instruction, among other secondary factors. When receiving a course or a lecture, a discourse that is internally coherent usually triggers a feeling of truthfulness and understanding, particularly when no much previous knowledge is available. However, the details of the discourse, and even the key arguments, are frequently forgotten after a short period of time. The impression of coherence remains, however, and maybe the conclusion.

Remembering the full contents of the discourse depends on former knowledge: if concepts can be linked to existing structures, thus activating such knowledge structures, and does not match exactly with them (e.g. adds more information or organises former knowledge in a different manner) the internal state can endure for a longer period, yet being lost after a given amount of time. Only if more reflection is devoted to the existing internal state (this is, the activated existing structures and the new materials) it becomes fixed in the long term memory (after being physically changed some parameters concerning brain states).

Therefore, it is a quite a common situation that the first impression after a given instructional experience does not predict the true usefulness of the learning (if any). The closer the time to the instructional situation the more intense the sensation of newness, coherence and usefulness. As time passes, and there is no further use of the information, most of the data vanishes, though something of the first impression still remains, frequently the sensation of being useful or interesting, alongside with a tag of the content.

2.1. Instruction and feeling of knowing

The feeling of knowing (FOK) is described as the sensation that someone has of having available information about a given subject in the long-term memory store. In a comparison, the FOK wold be similar to consult an index of a book, particularly an analytic index. If the reference exists, something concerning the topic is contained in the book. But there is only the reference of a page or some pages where the term appears. Searching in these pages sometimes provides plenty of information about the topic but, in other cases, it provides the disappointing experience of just getting a marginal reference to the content or the suggestion of further readings somewhere else. Normally, the analytical indexes of the books are made once the body of the book is fixed, so it seldom happens that a term that is cited in the index

has no correspondence with the contents. But this is not the case of human memory, which is in constant change and re-organisation. Information that was a part of our memory structures and was properly indexed can change or it can even be removed. Hence the possibility that a given index (feeling of knowing) points to nowhere or that it points to something different than the original target is not null.

Another distinctive property of the FOK indexes is that they are not only made of verbal tags (or whatever kind of representation is used, like images or abstractions). They normally include value or relevance estimations. Consequently, it is plausible that, after having received a course that provided the impression, in that moment, that the contents were useful and important, an entrance on the FOK index including the tag (the general content of the course) and the initial impressions would be created. If no further work is devoted to the true knowledge (the target of the FOK index) and it vanishes, the only thing that remains in the memory structures is just the first impression... pointing to nowhere.

The optimal situation takes place when instructed contents (or procedures) are connected to former knowledge (e.g. previous experiences) and subsequent reflection is enacted in different occasions. Each time the already existing or newly instructed knowledge is recovered, the brain suffers some changes that affect de simultaneous activation of the networks involved in these representations structure, therefore converting them, at the functional level, in a knowledge structure. Similarly, if the existing structure is scrutinized looking for weaknesses or contradictions, or attempts to use it to explain past or new practical situations are run through, the knowledge structure develops (new connections are defined, new materials can be added or wrong links and contents are removed) and the brain state associated with the central nodes of the knowledge structure becomes more and more solid. When appropriate meta-cognition is applied to the former processes, the FOK index is also updated both in the vector that points to the structure and the current importance of the structure.

The central idea, then, is that the FOK index may contain a tag of a given content and an estimation of its importance. Nevertheless it can point to a solid knowledge structure, to an intermediate structure that can be improved, to a poor structure or to nowhere. And the same tag and estimation of worthiness can exist for any of the knowledge structures enunciated in the last sentence.

The conspicuous consequence is that when someone is asked, for instance, if a given content has been of practical use for him or her, the easiest and quickest way to answer is acceding to the FOK index. Reaching the actual knowledge structure takes much more time and effort. Hence, in the cases that the FOK index is updated, the response will reflect the reality, but in many much cases, particularly in those where instructed contents have vanished, the index is not updated with the real structure. It may then just reflect the impression had at the instruction that the contents could be useful for practical purposes, though such contents are not present in any memory store.

On the contrary, when the requirement consist in using the knowledge, the index only serves as a way to activate such knowledge. And any real, complex product produced by a person involves whatever the knowledge that person has. Thereby, situated, complex products and behaviours can not be sustained on a mere index —they require the use of all the available knowledge. And this is the main reason why complex behavioural demonstrations and the analysis of complex products can bear witness to the existence of substantial changes in the knowledge structures. They involve true knowledge. On the contrary, subjective estimations of importance, usefulness or practicality are usually solved appealing to meta-

knowledge, which may not be updated and thus convey erroneous information.

2.2. Changes in knowledge structures

The changes that take place in learning can be explained at three different layers: first, what happens in the physical layer —normally the brain— where the threshold of activation or inhibition of a certain neuron-network is modified, maybe just in a part. That produces in the brain a different activation state or even changes the sequence of states of that network, in what is the physical manifestation of learning. Second, at the functional layer (i.e. the operations on representations performed by the brain) changes have to do with the amount of representations activated at once, the ease of recall or the stability of the representations in the memory store. All functional change relays on physical changes that took time and energy to occur. What drives these changes in the brain and subsequent modifications in the functionality is the *use* of the existing resources (e.g. representations, connections and reasoning sequences). What is not being used does not modify the brain states and, consequently, remains as it was when learning took place.

It does not exist in human brains such a thing as an universal mechanism devoted to reasoning. Indeed, the reasoning effectiveness depends on the properties of the knowledge structures that have been generated. A given person may have a sharp reasoning concerning one subject and a dull reasoning concerning another subject. The amount of connections and the truthfulness of these connections, as well as the quality of the representations make the difference. This point is particularly relevant, since it makes clear that first-class reasoning will not emerge spontaneously and can be applied to whatever information; it is rather a function of the elaboration of the underlying knowledge structure.

The third layer is the observable one: the behaviour. Human complex behaviour is mediated by how people represent and reason upon a given situation, thus receiving a strong influence from the functional level and, ultimately, from the physical layer. The details of what happens at the physical layer can be abstracted in most occasions in the same manner that one does not need to know most of the details of what is exactly happening in the engine to drive a car efficiently. However, it does not mean forgetting that the mechanisms, the physical mechanisms, do exist. In a similar manner that a car can run out of fuel, a brain may deplete its fuel, may need to have more neurotransmitters available or may need to clean the residuals of metabolic activity. It means that, although brains are incredibly powerful devices, they cannot be expanded to infinity. Learning always takes time and energy.

The main consequence is that the label “learning” includes a variety of situations that have very different demands on the physical bases and imply many different states of affairs at the functional and behavioural levels. For instant, remembering a point of information for a short period of time (say, some minutes) demands a small effort for most brains. The functionality (second layer) is very humble, since remembering does not involve reasoning, and the behaviour supported by it is also simple (e.g. declaring the remembered information). In the jargon of the functional layer (usually called *cognition* or *cognitive functions*) the information is “stored” in the “working memory (store)” and retrieved when necessary because some environmental demand exist. The more it is retrieved and the more varied the way this information is used, the higher is the fixation in the “long term memory (store)”. While the working memory has a limited capacity and duration, the long term memory is virtually infinite and stable. Thus, the observable behaviour of recalling the information and declaring it may be based on two different underlying states. It is certainly a proof of learning

but two different kinds of learning may have taken place, leading to two different physical states.

The example in the former paragraph was almost irrelevant, because the most interesting versions of learning usually involve plenty of changes in the long term memory and the coordination of many data-points (i.e. representations). Sophisticated behaviour, like integrating information, reasoning and producing a conclusion and acting it (that is what we usually call "being competent" on something) involves thousands of changes in the brain's networks and a lot of energy and time. Thus it does not happen fast and easy.

Table 1. Cognitive complexity of remembering, reasoning and behaving competently

	Remembering	Reasoning	Competence
Memory access	■	■	■
Representation with own resources		■	■
Connection with former knowledge		■	■
Knowledge re-organisation		■	■
Cognitive products		■	■
Decision making		■	■
Perceptual patterns			■
Response patterns			■
Response-oriented decision making			■

Table 1 summarises the large increase in cognitive operations that is needed in order to support higher cognitive functions, like reasoning, and complex behaviour. It should be remarked that many of the operations depend on the learners' characteristics, like the specific configuration of their brains (which provide a particular set of representational resources) or the existence of previous knowledge that can be connected with new materials. The difference of generating abstractions by oneself or "importing" these abstractions from someone else, for instance, is that self-generated abstractions are suitable for creating perceptual patterns while imported abstractions are not. Similarly, the organisation of knowledge and establishment of bonds will support good decision making if it is made by oneself and rather arbitrary decisions if it is "imported". Knowledge structures are the core of high level cognition.

An instructional action can provide the adequate pressures and opportunities to make these physical and functional changes happen. However, it must also be a sophisticated instructional action that previews many activities oriented to reasoning, linking existing and new materials or test the goodness of a mental model (a knowledge structure) against reality. Does it take a class? Maybe a few classes? No. It takes weeks or even months of work. The effectiveness of instructing complex learning resides on many operations that have to be done by the learner, who is the owner of the brain that changes when learning effectively takes place. Instruction can exert pressures in the appropriate direction, but cannot force or directly cause the activities that must be performed by the individual.

Many instructional actions (like courses) are interesting starting points of this kind of

learning process. This is particularly true when the professor is skillful enough to make learners activate adequate knowledge, previously consolidated, to expose gaps or imperfections in such knowledge, and to introduce the new elements that can be connected with former knowledge, suggesting better ways to organise it. Notwithstanding, this only provides an impression of coherence and usefulness. The suggested changes take place only after a wide dedication of each learner to reconsider by his or her own each of the steps, making his or her particular knowledge structures match and evolve. In this process, expert feedback becomes a central aid, though the bulk of the work is done by the student and a large part of this work takes place beyond the (frequently limited) duration of a course.

One obvious conclusion is that differences among students' capitalisation of instructional opportunities have a lot to do with how much extra time they devote to reflect on, inspect and test the new materials an models. Certainly they must be able of representing all the new materials and connect them to a sufficiently solid knowledge structure (e.g. previous experience); but, provided that these conditions are approximately met, eventual differences will depend on their ulterior individual work, not on the instruction received.

A typical instructional setting consists of an asymmetric situation where someone who masters an issue (because has devoted a lot of time to the scrutiny of that issue and has consequently developed a complex knowledge structure) provides cues about the changes to be done and the final states to be reached by a group of persons who are less knowledgeable. If the instructor is a true expert and the instruction is well designed, it can save years of erratic work and plenty of disappointments to the learners. But they must do their part of the deal. Something that took, maybe, thousands of work hours to the expert can be probably reduced to a smaller amount of work for the learners if they follow the wise indications provided by the instructor. But not to the short amount of hours spent in the course.

A double conclusion can be drawn from the points stated hitherto. On the one hand, the impact of instruction on students' knowledge, reasoning or behaviour can be certainly be modelled as a causal relationship, though instruction is not going the be the only cause and, in many occasions, neither the most important one. Student's dispositions and actual work will be more and more relevant the higher the complexity of the competencies to be learned.

On the other hand, expanding the duration of instructional support, not necessarily in the form of classes, is a better approach than increasing the number of courses or classes. Many courses or classes just mean that a lot of learning processes are being started simultaneously. The time and effort that are needed to develop all these learning processes is typically sacrificed in attending other classes (and starting new learning lines that will not be completed). What makes learning crystallise is redundance, using the same resource in different situations, connecting with personal experiences, creating and solving problems, finding gaps, trying alternative approaches. And receiving feedback.

The image of an expert making a copy of his or her knowledge available to the students, in the same way a computer file is copied on a pen-drive has absolutely nothing to do with instruction. Learning is always an individual process that has nothing magical. It is physical, and physical changes happen on the learner, in a part provoked by the environment and in another part provoked by his or her own actions.

2.3. Reasoning and behaving with knowledge

Knowledge structures are what we use for reasoning. And reasoning supports decision making and most sophisticated behaviours. There are many kinds of complex behaviours and

reasoning that also involve other kinds of learning. For instance, when action competencies are to be achieved, the construction of perceptual patterns (filters devoted to select relevant information and exclude noisy data) and action skills (response organisation patterns) are two classes of learning to be developed alongside the development of the central information structure. Although necessary, perceptual patterns and action skills can be considered as complementary, and they are typically constructed in the natural setting were behaviour takes place. The central knowledge, the one that represents the object that is manipulated when reasoning, is the structure that is most susceptible to be instructed in order to start its development. The main part of reasoning and understanding will depend on how well-developed such structure is.

As it was stated in the former section, instruction usually initiates the process of transforming a structure of knowledge and improve it. But it is individual work and effort what eventually perfects it. The important point is that, since behaviour and decision making is grounded on knowledge structures, existing knowledge can be inferred from the way someone behaves, particularly if there are steady products associated to the behaviour.

There are, however, some reasons that may interfere in the use of a given knowledge structure, even when there is a well elaborated structure. In general, such a structure will provide behavioural advantages thus, if no interfering conditions exist, the normal situation would be implementing the knowledge and derived reasoning in the behavioural manifestations. It takes some time, particularly when a former behaviour structure should be replaced. Although the new behaviour structure should demonstrate to be more efficient, because is grounded on a better representation of the issue, existing behavioural structures are usually more automatised and can be triggered with ease and low energy-consumption. Therefore, replacing existing behavioural structures involves two operations: the effortful inhibition of existing mechanisms and the conscious use of the new behaviour, which has not been automatised yet. After some time doing this double action, inhibiting the former behaviour pattern becomes less effortful and the new behaviour becomes more and more automatic. Eventually, the old behaviour is virtually replaced and the new one is triggered instead. The former mechanism does not disappear —it can take years of not being used if it was well fixed— but a shift has happened in the sense that it has to be activated consciously, while the new response pattern is activated automatically. If there were no existing previous behaviour, the process would be far more easy and swift, since all the energy could be devoted to consolidate the new behavioural pattern from scratch.

Nevertheless, the most pervasive source that may hinder the implantation of new and more efficient behaviour has to do with social pressures and social conformity. When one is not enough self-confident, reproducing the patterns of behaviour that are more widespread is a good way to avoid confrontations, even though one is aware that these patterns are imperfect or erroneous. Young professionals, for instance, may have a lot of trouble if they differ significantly from the mainstream ways of behaving. That may involve challenging more experienced professionals, who have a wider network of contacts and can be more influential, thus losing possible allies (or having more enemies). Hence, if no explicit support is provided “from above”, furnishing the prestige of the new way of acting, many people may feel some degree of hostility that will block the use of a well elaborated knowledge structure in their actual behaviour.

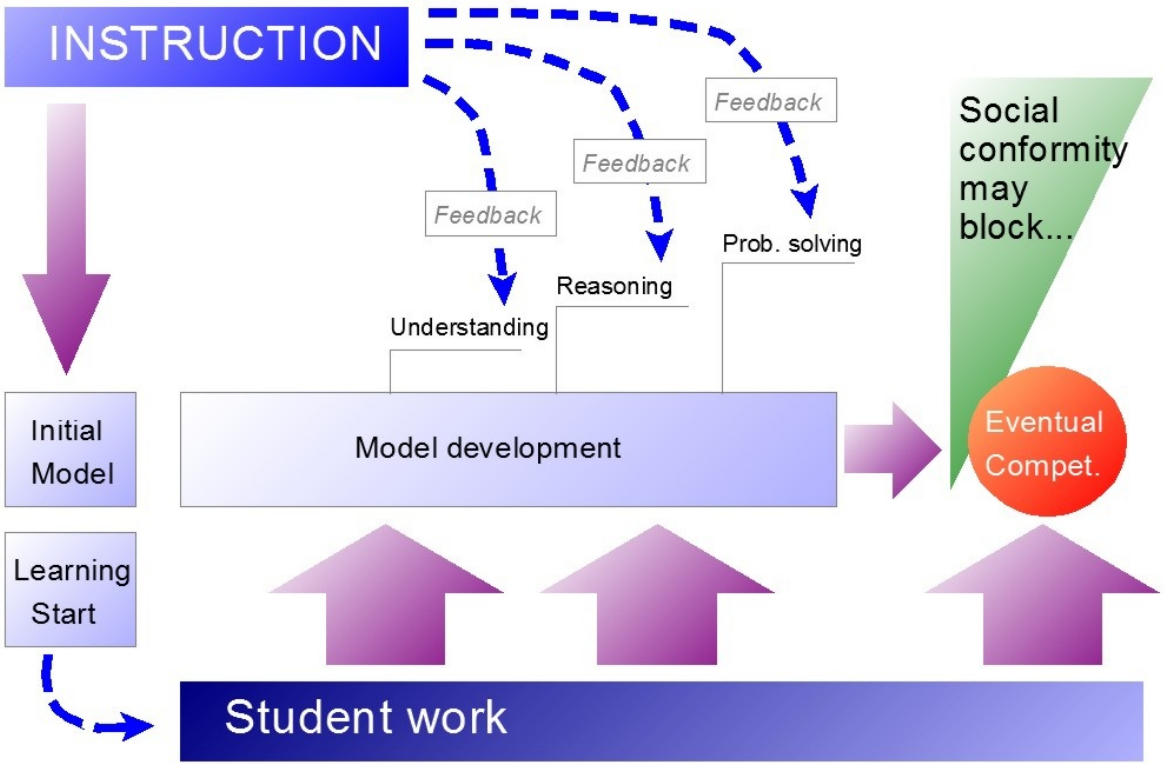

Figure 1. Instruction and student work contributions

Figure 1 depicts the relative influence of instruction and the independent work of each student in the completion of the knowledge structure that models the professional field.

The eventual competence, which is a behavioural dimension, may be blocked by social conformity effects. In any event, further instruction can be added in order to optimise the competence and to neutralise the effects of social conformity disruptions.

3. Overall design

The situation described in the previous sections indicates that it is rather difficult that instruction, by itself, has a large effect on instructed persons' behaviour. It is nevertheless important to state that, although the effect might be small, this effect is the outmost goal of an instruction process addressed to professionals. It is not a matter of producing changes in the knowledge structures alone —although it is a worthy achievement— but also to succeed in the modification of the behaviour of those that received instruction.

Since the instruction of judges and prosecutors fits perfectly in the conditions described above (including a probable pressure addressed to new ways of behaving) the expectation of behavioural change cannot be high. Notwithstanding, it can neither be null. Hence, it is a reasonable hypothesis expecting the professional behaviour of those that have received instruction from the Judicial Academy to be different than the behaviour of those that have not. And this is going to be the hypothesis where the design is founded: the comparison of objective behaviour of those professionals instructed with those that were not instructed. The behaviour is going to be analysed in two connected dimensions: first, its general quality and coherence; and, second, the presence and use of contents specifically taught in the Judicial Academy instruction.

3.1. Strong and week proofs of instructional effectiveness

The instruction addressed to professionals should always be aimed to achieve stable changes in the way these professionals behave. Other kinds of changes may ultimately have some effect on their behaviour, though it cannot be taken for sure. For example, they can be glad about the instruction received, evaluate the subjects taught as significant, or even modify their representations and knowledge structures. But if the instruction does not modify their professional way of behaving, not much can be said about its effectiveness.

It can certainly be stated that, in general, instructional effects can be evaluated from a large range of indicators. Some of them are very weak proofs of usefulness and some others, even if the traces of the effects are small, constitute a solid proof. Ordering them from weaker to stronger, the following indicators can be used:

(1) Satisfaction expressed immediately after the courses;
(2) Feeling of knowing immediately after taking the courses:
(3) Knowledge demonstrated in exams or equivalent tests connected to the courses;
(4) Feeling of knowing after one year or more the courses have been taken;
(5) Feeling of knowing after two years or more the courses have been taken;
(6) Actual knowledge present in their knowledge structures after one year or more.
(7) Behaviour of the trainees after one year or more.

Satisfaction after the courses is strongly influenced by all the conditions that were present when instruction took place. It mixes a set of first impressions with immediate experiences (like the interactions with the classmates). It is a useful index to detect things that did not work, though a bad predictor of learning. The feeling of knowing becomes a more solid predictor as time passes but if, and only if, there has been some independent work from the learner side and the index has been updated accordingly. If not, the index reflex the feeling immediately after the course was taken while the leftover contents may be fewer or none.

Indicators 6 and 7 are the more valid ones, since they imply a direct access to knowledge and its use. In fact, changes in behaviour necessarily involve changes in knowledge, particularly those that are action-oriented and not declarative.

3.2. Expected changes

There are two different scenarios that can take place: the student do not devote more time to reflection on the contents and the development of the knowledge structure associated to them; or the learner starts thinking about the contents and developing a personal model. The first situation will probably lead to a gradual loss of the information stored in the working memory, thus most of the contents will vanish after few days. Those contents that accessed the long term memory store, if they are no further elaborated, will become invisible —unaccessible, indeed— after a couple of weeks. The feeling of knowing index, however, will remain as it was immediately after the course. Hence, the impressions will prevail, although associated to almost nothing. If these impressions are surveyed the answer will not reflect the preserved knowledge, which will presumably be less.

When the student has kept thinking on the courses' contents, the expected situation varies significantly. First, the contents are usually transformed from the original representations used by the professor to a set of representations that are comfortable for the learners (that is, that better suit their brains' representational capabilities). In order to convey information, verbal- or image-coding is typically used. This codification, notwithstanding, might not fully represent the conveyed object. For instance, the "feelings of a victim of a crime" is just a verbal label that refers to a set of physical and emotional sensations. A proper representation of the content involves transforming the words in true feelings, thus making an emotional-coding instead of a verbal-coding. The meaning of the contents changes —normally improves— when re-coded in the proper manner and, typically, they become easily connected with former knowledge, either instructional or experiential. Hence, the taught contents evolve in the sense of gaining significance for the learner and being re-connected with previous knowledge structures, which is a great way to consolidate understanding.

In doing these conversions and connections there may be errors or missing elements. At this point, having feedback from the instructor (provided that the instructor is a true expert) is of great help. If possible mistakes are detected and corrected, and lacking elements are integrated in the structure, the core of knowledge is solid and can grow up for years. On the contrary, if errors are present or significant parts of the contents are missing, the structure of knowledge gets birth contaminated, and will probably grow up making the errors greater.

A further step in knowledge development takes place if the learner starts checking over the new structure looking for gaps or inconsistences, and generates cognitive products (like abstractions) of it. New experiences are usually connected with the growing structure too, producing an accumulative set of adjustments and improvements. The main consequence is that the knowledge structure not only grows but it also becomes more useful, since it is connected with practical experiences, and it better supports reasoning (through abstraction, for instance, which is great for logic). If meta-cognition is applied (something quite probable after such an elaboration of the knowledge structure), the feeling of knowing index is updated and upgraded with the new potentialities (such as practical uses, reasoning) associated with the evolved knowledge structure.

Once this level of perfection is achieved, behaviour can take advantage of the

knowledge and reasoning provided by the mental model (or, with equivalent meaning, knowledge structure). Problem solving becomes improved and, if perceptual patterns and skills are properly consolidated, true competence becomes a fact. At this point, limitations could emerge from social pressures or social conformity, thus inhibiting the competent behaviour and replacing it with a more conventional one.

From the evaluation point of view, the final behaviour (or competence) is the utmost proof of a proper expansion of the knowledge initiated by instruction. Nevertheless, when social pressures are detected, a direct evaluation of the existing knowledge structure can serve as an alternative way, at least while trying to reduce the pressures that block behaviour.

Changes in feeling of knowing

When learners are asked about the importance or usefulness of what they have learned, they commonly search information in the meta-cognitive store (i.e. the feeling of knowing index). Thus, depending on how much development their knowledge structures have had, the index will reflect the state of the art or it will rather be a misguiding record of the first impressions.

In terms of objective evaluation, this kind of survey is not the best choice. Even so, sometimes is needed to have over-optimistic data, based on the opinions of the trainees, to (apparently) justify the appropriateness of an instruction program. This data is still useful to distinguish, for instance, the courses that made a good impression from those that did not. The central problem is that appealing to the feeling of knowing mixes information from those trainees that kept the first impression with information coming from those that improved the structure of knowledge and updated the index. As it is impossible to determine, without accessing to the knowledge structures, how much learners are in each situation, responses cannot be balanced.

In general, still, people who have elaborated a given content should be expected to provide a feeling of knowing response that has a true correspondence with actual knowledge and behaviour. And that is a more economic way to gather information, compared with costly accesses to knowledge structures and behaviour.

Changes in knowledge structures

These changes correspond to the transformations that are steadily fixed in the individual's long-term memory. They are a central part of the true learning achieved, probably initiated by the course and consolidated through individual work. Any evaluation design aimed to this knowledge is, by far, more valid than most of the indirect measures (like satisfaction or FOK).

It should be kept in mind that knowledge structures that model an applied situation, like a professional setting, may include plenty of procedural knowledge, which is not so easy to make explicit by verbal means. For instance, facts can be categorised in blurry classes (like "cases difficult do defend") where the label gives no much cues to an external observer, though the class may have perfect sense in the framework of the mental model of who has created it. Situations like this make verbal approaches to knowledge elicitation of limited use. Opposed to the verbal elicitation, behavioural analysis usually shed much more light on knowledge organisation (for instance, grouping a set of cases in the same strategy of defence). Abstractions spontaneously generated by each person serve these kind of purposes, like associating some characteristics of a situation with a typical response to it, but they are

seldom labelled in a precise manner.

Some professional activities, notwithstanding, have strong demands in making explicit many of these classes and criteria to assign a case to any of them. A teacher or a journalist need to do so for communicating, as well as judges and prosecutors must produce legal documents (sentences, judgments, claims) that clearly reproduce their reasoning and argumentation. Consequently, the written materials can be considered a proper manifestation of their behaviour that makes explicit most of the knowledge they have.

Changes in reasoning and behaving

In a normal human cognitive functioning, knowledge is used for reasoning, decision making and adjusting behaviour to meet environmental demands. Situations where knowledge just have to be retrieved are artificial and, in a sense, deviant. Most applied settings, like professional ones, have a small declarative space and a large practical dimension. And that makes the elaboration of knowledge to develop very differently when it is aimed to application or declaration. For instance, experts in a given field that also exert as a professors have to transform most of their practical knowledge in a format that can be conveyed to (and more or less understood by) their pupils. It normally involves adding a further layer to their knowledge structures with the declarative representation. This is usually a harsh task because action-oriented knowledge can prescind from many labels (typically verbal) or use fuzzy probabilistic categories instead of strict logical ones.

At the time the new layer has been properly constructed, it can be considered an improvement in the knowledge structure and usually also improves understanding and reasoning. That is the reason why many professionals of teaching declare that teaching has also been useful for them to better understand the topics taught. In any case, the central idea is that cognition has not evolved to satisfy academic demands. They are just a particular setting, somewhat anti-natural, where recall to merely declare plays an outstanding role. Recalling for being used in reasoning and decision making is the natural way to use knowledge. Furthermore, reasoning is normally implemented by following the links established in the knowledge structure. Sometimes it provokes some modifications of the structure itself, like when a new connection is envisioned or an error is detected at the light of new situated data.

The central idea, hence, is that knowledge and reasoning are intrinsically bonded, so any change in the knowledge structures is going to modify the way one reasons, and reasoning may led, in some occasions, to the adjustment of the knowledge structure. From an evaluation point of view, that also indicates that the way a persons reasons can tell a lot of information about the underlying knowledge structure. Conversely, behaviour is supported by reasoning and decision making, and therefore by knowledge.

Still, social life involves a sophisticated set of rules and procedures that may interfere or even contradict the expected behaviour according knowledge and reasoning. For instance, courtesy rules make us treat kindly both people who is positively represented and other people who may not be so positively represented. In other words, social behaviour can be rather disconnected from the underlying knowledge and reasoning. In social terms, knowledge and thoughts are seldom used; social life is based on behaviour. For instance, what a person thinks is rarely a problem, except when that person declares it. Consequently, it is not unfrequent to find a certain degree of dissociation between thought and behaviour. Power is not always linked to reason and that makes many persons to be very cautious about how they behave, despite the logical implications of the knowledge they manage.

Again, thence, actual behaviour may reflect few influences of the underlying knowledge. It is nevertheless true that, when the knowledge structure has been thoroughly elaborated, it does effect behaviour, although the main modulation of such behaviour comes from other (social) sources. It implies that evidences of the changes in knowledge and thought should be observed, even though their magnitude could probably be small if strong social pressures exist.

3.3. The control group

Since small differences are expected, the estimation of the gain or variation that has taken place in the trainees is not as much important as demonstrating that they do behave differently than those that did not took the courses. If no differences existed, it would imply that the courses had no effect on behaviour and probably not much on knowledge structures. Evaluating the knowledge structures directly could determine whether knowledge exists but is not implemented in behaviour, or it rather does not differ from the knowledge used by untrained professionals. The last case would imply that training had no practical effect at all.

In any case, the design is simple: the expectation of instruction to modify professional work-performance can be tested by comparing a group of trained professionals with an equivalent group of untrained professionals.

The parallel groups design

Parallel groups means that the groups only differ in the independent variable studied. This is to say that other relevant variables are balanced. In this cases, experience, age interval and geographical placement of the workplace (paying special attention to the type of juridical cases that are typical) are the main variables to be balanced. These variables include both general cultural conditions (age and geographical setting) as well as professional experience.

Instructed vs. not instructed

This is the qualifying variable to assign cases to any of the groups. The procedure should consist of selecting participants that are currently working in the administration of justice and have taken the courses at the Judicial School. Once the list is completed, a parallel list of participants that have not been instructed is to be defined. Each participant in one lists is matched with a participant in the other list with the three balancing variables mentioned in the previous section having equivalent values. And they only (or mainly) differ in the factor studied: having received instruction by the Judicial Academy or not.

The object of the measurement is a situated behavioural product: the legal documents produced by the professionals. These documents satisfy the requirements of the most valid evidences: they are situated (they make sense in natural settings and are not measurement artifacts), they are complex products (so every available cognitive resource may be involved in their production) and they are stable (allowing a deep scrutiny and replication).

The prospect is that the eventual effects of the training received at the Judicial School courses and its ulterior development by the trainees should be expected to be cast in foundations, argumentation and reasoning included in such documents, which would differ depending on the group that has written them. Hence, it is possible to establish a protocol of dimensions to look at, and perform an analysis of the documents by law experts. This analysis

would be “double blind”, in the sense that reviewers would not know whether the evaluated document belongs to someone who took the courses or not. They would simply verify and quantify the points in the evaluation protocol. The “double” means that the evaluation team would also have no intervention in the committee that applies the protocol. The quantification of the experts committee would go back to the general evaluation team and they would associate each score to each person and, therefore, to each group. After that, statistical tests would be applied to compare the two groups and would uncover, if there existed, the eventual differences.

Expected differences

The expectation, according to what has been stated in present document, is that the group that has been instructed will score higher in the legal foundation, use of arguments and reasoning procedures that were taught in the Judicial Academy.

It does not mean that they get the maximum score according to the protocol, but scoring higher will mean that instruction has had an effect on their knowledge structures and on their professional behaviour as well. As it has been said, important obstacles related to the use of the changes in knowledge triggered by the instruction can reasonably be considered to exist. Thence, the finding of small, though statistically significant, differences in such an hostile conditions would be a solid proof of the effectiveness of the courses taught.

3.4. Consequences and decision making

Setting the expectation apart, there are different outcomes that should be considered and decisions to be taken. On the one hand, three possible outcomes can result from the statistical comparison of the parallel groups:

(A) The instructed perform *higher* than the non instructed. That will prove that the instruction task developed by the Judicial Academy is effective. The intensity of the effect will be considered below.

(B) The instructed perform *similarly* to the non instructed. There are no statistically significant differences. In this case there are no proofs that the professional behaviour of the trainees is modified by the training. It would be recommendable to check whether there exist differences in the knowledge structures. In any event, it would mean that the training has not been effective. Differences in knowledge structures would suggest that final implementation should be stressed in the training procedure. If no difference was found, the whole instruction procedure should be reconsidered.

(C) The instructed perform *worse* than the not instructed. In this case results would indicate that instruction has a counter-productive effect on trainees. Instruction should be deeply reconsidered.

If the situation (A) were the result, then it would be worth considering the percentage of fulfilment of the evaluation protocol (i.e. the intensity or magnitude of use). The difference between the average scores of the instructed and non-instructed groups could be a reasonable estimation of the instructional effect. As a fictitious illustration, be the maximum score in the

evaluation protocol 50 points; suppose that the instructed group has an average score of 23 points and the non instructed group and average score of 14 points. Then the following inferences can be drawn:

a) A 28% (14/50) of the performance can be achieved without instruction, this is through other means like the press, general culture or basic professional instruction.

b) The contribution of the instruction can be estimated as a 18%, since the average performance of the instructed group was 46%, but 28% was achievable by non-instructional means.

c) In any event, the performance of the instructed group is slightly inferior to the half of the optimal.

Overall, the three inferences stated indicate that the instructional process has been useful though still far away from the optimal effect. It has already been mentioned that, when assessing situated behaviour, many other sources of variation (particularly social ones) must be taken into consideration. In this context, results can be considered as satisfactory and will probably imply a mid- to long-term increase, as more and more professionals share the instructed principles and contents.

It does not contradict, however, that stressing the implementation of the instructed contents could be a desirable improvement from the Judicial Academy side. For instance, a complementary course for those that received current instruction could be programmed.

Consider now a different fictitious example: be the maximum score in the evaluation protocol 50 points, like in the former simulation; suppose that the instructed group has an average score of 47 points and the non instructed group and average score of 42 points. The following inferences can be drawn:

a) A 84% (42/50) of the performance can be achieved without instruction.

b) The contribution of the instruction can be estimated as a 10%, since the average performance of the instructed group was 94%, but a 84% was achievable by non-instructional means.

c) The performance of both groups is very high, being the performance of the instructed group almost optimal.

In this example, although demonstrates that the instruction has an effect it would be reasonable considering whether this effect is strong enough to justify the inversion in the course. The “spontaneous” level of achievement is really high (84%) and probably the improvement due to instruction would not surpass a cost-benefit analysis.

4. Evaluation criteria

The committee that will judge the written documents produced by the professionals in the analysed groups must be provided with a set of objective criteria to apply. The more detailed and exhaustive the criteria, the more reliable the judgements are going to be. Anyway, criteria cannot illustrate with maximal precision what to look for, in the sense of finding a given expression, for instance. The decision of whether an expression satisfy the criteria or not depend on the expert knowledge and experience of the members of the committee.

The set of criteria can take the form of an overall evaluation of the instruction generated by the Judicial Academy, in which case it would focus on general properties of the documents, like a solid legal foundation or coherence in the argumentation and conclusions. But it can also assess specific-subject instruction, focussing on the presence of particular arguments, foundations or procedures that were taught in a particular course or set of courses. And the second approach is perfectly compatible with the first, including general properties of the document.

4.1. What to look at

There are a group of indicators that can be used to organise the protocol. They are strongly recommended to be included in whatever criteria. The specific concretion, importance and quantitative value of each will be determined by the committee in each case. Nevertheless, similar scores should be assigned to each category (for instance 10 points to all of the elements; or 10, 12, 8, 12, 12 to stress the importance in a given case; et cetera). The points that are assigned are arbitrary in its magnitude, but should correspond to the minimum value that permits the discrimination of the existing (or possible) levels of performance in the legal documents. It is also a good idea to have a model of the "perfect" document, as well as illustrations of the intermediate scores.

The central indicators are as follows:

Legal foundations

The references to laws that should necessarily be done. In this case, the score assigned to each reference should depend on its importance. For instance, a central reference could score 5 points, while secondary references could score 3 points and marginal references 1 point.

Weighting the foundations

Citing a legal regulations is not enough. If a regulation is important, it should appear in the argumentation and conclusions with the proper weight, according to its importance. Scoring principles before mentioned also apply.

Use of instructed contents and procedures

Specific contents (concepts, procedures) taught in the instruction should be present in the documents. A list of such concepts and procedure should be provided by the persons who designed the course. They can have different scores, according to their importance, or simply be counted (how many concepts and procedures are present or how many times they are

used).

Coherent argumentation

Correct use of the knowledge involved. The sense of the concepts and procedures used should be the correct one. Errors, omissions and misunderstandings should be scored negatively, while correct uses would be scored positively.

Expert knowledge

Expert knowledge consist in going beyond the instructed contents. It is a solid demonstration that the person has reflected on them and elaborated personal conclusions.

4.2. The role of a renowned law expert

It is a central point for the objectivity of the evaluation committee that the members that compose it should be prestigious and respected by any of the evaluated persons. Then it would be highly recommendable that the chairman was a renowned jurist whose prestige comes from professional, politically-independent merits. It would also be an extra guarantee that this person was not Serbian (could be Croatian, for instance, in order to avoid language-related difficulties). The role of that person should be to decide the criteria to be applied and supervise the evaluations made by the members of the committee. He or she can also belong or have belonged to International Juries or International Judicial Organisations (like the Human Rights Tribunal).

The higher the prestige and independence of this person, the stronger the credibility of the evaluation results. The remaining members should also be as much prestigious and independent as possible, though they would not be the visible face of the committee.

4.3. The evaluation committee

The way to proceed of the evaluation committee should consist in dividing the documents to be evaluated among the members (may be excluding the chairman) and proceed to a first evaluation according to the criteria previously established. They would not know whether any of the documents have been produced by a professional instructed by the Judicial Academy or not.

After the first evaluation is completed, the cases should be discussed by all the committee and, if agreed, a final mark assigned. Alternatively, the chairman can supervise all the documents and scores assigned to them according to the established protocol, and introduce the justified modifications he or she considers opportune.