Operationalisation

Operationalisation is providing a precise, concrete definition of a variable in such a way that it can be measured.

How Questions Emerge From Topic Areas Or 'Problems'

Research questions can take time to develop. While in many ways they are the starting point for the focused research, they can take weeks or months to develop. In some fields and on some projects, the whole aim might be, over several years, to work towards a research question!

Let us assume that you want to get to your research question within weeks or months. There are two ways to go about the task. One is to work hard and fast, early in the project, to generate and refine your research question. The other is to let the research question emerge from the literature review.

In the first half of 2001, I co-ordinated a project team which was trying to generate a research question in the area of information communication technology (ICT) and literacy learning. We had agreed that this was the broad area in which we wanted to undertake a systematic literature review, but we also knew that the two fields were so large in themselves that the project would be unmanageable if we didn't formulate the research question sharply enough.

Our first step was partly determined by the quasi-scientific world of the systematic review (a literature review which attempts, systematically, to cover a wide range of the literature in as unbiased a way as possible).

We were working within a paradigm which expected to find causal relations between different phenomena. In our case, the first move was to determine that we were interested in the effect or impact or influence of ICT on literacy learning. Initially, we were hoping to maintain a two-way relationship between ICT and literacy learning, i.e., what is the influence of ICT on literacy learning and in turn, what is the influence of literacy learning on ICT? We understood at the time - and were later to understand more fully - that the relationship between ICT and literacy learning is symbiotic, and that a one-way causal relationship would only paint part of the picture in this field. Nevertheless, we decided, on the advice of teachers and researchers, to stick to one aspect of the symbiotic relationship: what is the influence/effect/impact of ICT on literacy learning?

As soon as you are beginning to focus on influence or impact or effect, you need to decide how sharp that focus will be - or, more accurately, how tight the aperture will be of your study. *Influence* suggests a wide aperture; *impact* a smaller one; and *effect* the tightest of the three. Because our attention was in a field - education and learning - with a large number of variables and factors at play, we wanted to keep the aperture open a little more widely that would have been suggested by the use of 'effect', and so we decided on 'impact'. Here's how we justified that decision:

> We have chosen 'impact' rather than 'effect' as we wish to determine the broader aspects of the influence of ICT on learning in English (and vice versa) rather than merely attempt to measure effect. The term 'impact' allows us to examine strategies and processes as well as outcomes; we use it to include 'effect' but also to refer to wider (not necessarily causal) influences, e.g., on curriculum or policy.

To choose 'effect' would have meant that we would need a very different kind of study. To

measure effect it is almost always necessary to set up a control group and experimental group. The effect of an intervention is measured in the experimental group by comparing it with the effect (or lack of it) in the control group. The aim of the experiment is to determine the effect an intervention has; and its reliability is a measure of whether the same effect would have been observed if the experiment had been undertaken with another group of equally randomized subjects or respondents.

Our choice of 'impact' rather than 'effect' was critical in our study, because it allowed us to look at a wider range of study types than the randomized controlled trial. In other words, we were able to look at most study types.

So far, then, the research question had reached the state of 'What is the impact of ICT on literacy learning?'

Before we go on to delimit and contain the question further, let us pause and consider what we mean by two key terms in the question: *ICT* and *literacy.* Most research projects define key terms so that it is clear what is meant, in particular, by these terms in the course of the research and in its reporting. Our project was no exception:

> lCT stands for 'information and communication technologies'. In the proposed systematic review, we limit our-lives to networked technologies with a multimodal interface, i.e., networked and stand-alone computers, mobile phones with the capacity for a range of types of communication, and other technologies which allow multimodal and interactive communication. In terms of the use of ICT in subject English (and other subjects in the curriculum) , the National Curriculum for England suggests that '... pupils should be given opportunities to apply and develop their ICT capability through the use of ICT tools to support their learning' and they 'should be given opportunities to support their work by being taught to (a) find out things from a variety of sources, selecting and synthesizing information to meet needs and developing an ability to question its accuracy, bias and plausibility, (b) develop their ideas using ICT tools to amend and refine their work and enhance its quality and accuracy, (c) exchange and share information, both directly and through electronic media and (d) review, modify and evaluate their work, reflecting critically on its quality; as it progresses' (DIEE/QCA, 1999 (English) p. 52).
>
> We have chosen 'literacy' for a number of reasons: first, to delimit the field of enquiry to reading and writing; second, to distinguish the learning literacy from the subject English as taught in the National Curriculum for England [the National Curriculum was established in England and Wales in the early 1990s and is statutory for the 5-16 year old curriculum in state schools]; third, because as a term (especially in its pluralistic sense of 'literacies') it is both narrowly definable and open to wider interpretation; and fourth, because it allows us to review research that takes place outside formal education, e.g. in homes and other communities in which young people operate. In the proposed study, we will focus on reading and writing (in the broadest senses of those terms). Such delimiting of the focus of our review does not mean that the results of the review will not be relevant to the teaching of literacy in other countries; nor that we will limit ourselves to research undertaken in England. On the contrary, our net is spread wide.

These initial discussions of definitions were refined into shorter, working definitions for the purposes

of the research process:

> ICT stands for 'information and communication technologies', networked technologies with a multimodal interface, i.e., networked and stand-alone computers, mobile phones with the capacity for a range of types of communication, and other technologies which allow multimodal and interactive communication.
>
> 'Literacy' can be defined narrowly, as the ability to understand and create written language. But firstly, the scope can be expanded so that written language becomes written language and graphical or pictorial representation. Secondly, the skill can be treated as social, rather than psychological; in this view literacy is the ability to operate a series of social or cultural representations. Since sets of expectations and norms differ depending on the situation, the social view of literacy entails a number of different 'literacies'.

The point here is that definitions themselves might have to be refined during the course of determining the research question.

At this point, we are not much further on in the shaping of the research question, but we do have a very sharp sense of what the core of the research question means: 'What is the impact of ICT on literacy learning?'

You will have noticed that the word 'learning' is still undefined. This is because we have taken 'literacy learning' as a compound concept. However, the concept needs further definition. We achieved such refinement by adding elements to the research question. First, we added the fact that we were limiting the study to literacy learning *in English* - and by 'English', we meant the English language as spoken, listened to, read and written and as part of multimodal communication either as a mother tongue or as an additional language. Our proposal or protocol read as follows:

> By 'English-speaking countries' we mean any country or state which uses English as a first language or second/ third language, especially where in the latter case English is not only the language of research publication, but also a language taught in the curriculum between 5 and 16, e.g., Hong Kong, Singapore

Essentially, then, we were going to look at what is commonly known as 'English as a Second Language (ESL) ' or 'English as an Additional Language (EAL)' rather than at 'English as a Foreign Language (EFL)'.

Our research question now looked like this 'What is the impact of ICT on literacy learning in English?'. We had thought of further defining 'English' by delimiting it to the subject English as taught in the National Curriculum for England and Wales. However, the use of 'English' in that context clashed with our broader definitions of literacies', especially with regard to multimodal and visual (still and moving image) literacies, and thus was thought to be unsuitable; furthermore, a definition of `English' based on curricula in England and Wales would not have suited the international range of reference we had in mind.

Finally, we wished to limit the particular study to the compulsory years of schooling in most countries, so the final research question became 'What is the impact of ICT on literacy learning in English, 5-16?'.

The advantage of such a sharply honed question is that it helps in defining the scope of the literature review, as well as providing an answerable question that will determine the structure and development of the research project. Our systematic review required us to go one stage further, and develop inclusion criteria to determine the direction, nature and quality of the literature review. To be included in our review, studies would have to:

- be a systematic review, an intervention (outcome or process evaluation) or a non-intervention;
- have as their main focus ICT applications to literacy development;
- focus on literacy learning and teaching in schools and/or homes;
- be about the impact of ICT on literacy development;
- be published in English, in the period 1990-2001;
- look at literacy and ICT in English-speaking countries; be a completed study;
- be studies whose participants/study population includes children at ages 5-16;
- not be opinion pieces.

You can see from the inclusion criteria that we have added further constraints: on time (1990-2001), that the work be published in English, be completed studies and not be opinion pieces. Even with such meticulous refining of the research question, our searches of a number of electronic databases and hand-searches of a number of journals in the field revealed nearly 2000 articles, chapters and books: more than enough, as an initial trawl, for a Masters or Doctoral thesis.

During the course of the study, we discovered that to answer the research question, 'What is the impact of ICT on literacy learning in English, 5-16?' involved a number of subsidiary questions. For example, ICT's impact on literacy can be manifested in word-processing and composition, via networked technologies (email, the Internet), on reading, on dyslexia and so on. In practical terms, we decided to complete one of these subsidiary areas before embarking on the answering of the overarching research question. In this example, then, a subsidiary question, 'What is the impact of *networked* ICT on literacy learning in English, 5-16?' became a contributory research question - a contribution to answering the overall question with which we embarked on the research.

(Andrews, 2003, pp. 9-15)

Operationalization

The *operationalization* of variables is a researcher's chance to explain how each variable is being defined with respect to the construct in question. Such an *operational definition* should take a variable out of the realm of theory and plant it squarely in concrete reality. Basically, it must be a definition that is based on observable, testable, or quantifiable characteristics. Moreover, an operational definition must be unique, or exclusive; the definition must *not* also fit other possible constructs.

Let us return to the construct labeled language proficiency. This construct is hard to define, yet we all have a feel for what it is. Part of the problem is that it is an abstraction and is not directly observable because it has no tangible referent in reality. Of course, most researchers in language studies would probably limit the scope of this construct by using the compounding and adjectival processes that are so productive in English. Thus, we would probably encounter a much more narrowly labeled abstraction in any real study, such as a construct called "overall proficiency in English as a foreign language**".** One step, then, in looking at a

researcher's operational definition of variables is to evaluate whether the construct has been labeled with adequate precision, both theoretically and practically.

Although the construct in this example may now seem narrower in scope, it should be clear to any language teacher that "overall proficiency in English as a foreign language" is still a broad abstraction. To bring this construct down to earth and form a variable, the investigator might choose to define it as follows: overall proficiency in English as a foreign language as measured by the *Test of English as a Foreign Language* (TOEFL). We now have an observable and quantifiable definition of what the researcher means by this construct. The author has described, in concrete terms, what is meant by the variable. This is an operational definition of the variable.

It is the reader's responsibility to make sure that the variable, as operationalized, makes sense - that it logically represents the construct involved. To this end, several questions might be posed:

1. Is the construct labeled with sufficient precision?
2. Does the operational definition of the variable adequately describe the characteristics of the construct in question? (Is the TOEFL an adequate test of overall proficiency in English as a foreign language?)
3. Could the definition describe any other constructs? (Does the TOEFL test other things?)

The importance of thinking about a study at this level should not be underestimated. The results of a study cannot be logical if the variables are defined poorly or erroneously. The variables are, after all, the focus and center of any study.

(Brown, 1988, pp. 8-9)

Operationalizing variables

Although all types of constructs may be analyzed, the one operationalized as the independent variable is normally a grouping factor, or nominal variable. The dependent variable, in contrast, will typically be on an interval or ordinal scale. The choice of the correct statistic will depend on (1) the number of groups involved, (2) the types of scales used, and (3) the sizes of the samples in each group. And, of course, the results of such studies will be no more reliable, valid, or meaningful than the tests, questionnaires, or definitions of categories that are provided to operationalize the constructs as variables. Hence, it is, once again, important to consider the variables involved in such a study.

For instance, the hypothetical study in Example 11.1 contains one dependent variable: overall proficiency in French, as reflected by scores on the French proficiency test, an interval scale. There is also one independent variable in the study: whether students were in the treatment group (extra laboratory training) or in the control group (no extra training). This variable would be a nominal scale.

Are you satisfied with the way these variables have been operationalized? Consider just the independent variable for a moment. How can you be sure that students in the control group did not sneak into the laboratory late at night or have their friends in the experimental group smuggle cassette tapes out? If either of these cases were true, the control group's scores might be higher than those of the experimental group but for reasons that are out of the hands of the researcher. Surely, a study of this size would be the talk of the French department. What effect might that have on the internal and external validity of such a study? And why should you accept the French proficiency test at face value? The author says nothing about its reliability or validity as a test of French proficiency for this group, It is

just such imprecision in the operationalization and design of research that you should be alert to in reading actual studies.

(Brown, 1988, pp., 156-7)

Operationalizing variables

Your ability to obtain evidence for research questions will remain at a stand-still until you are able to translate abstract, vague, squishy constructs like *charismatic leadership* into something specific, concrete, and ultimately, observable and therefore measurable. For example, the conceptual definition of *workaholic* could be operationalized as *'someone who works more than x hours per week'.* It could also be defined through seeking the employee's degree of agreement/disagreement to attitude statements such as *'I hate not being able to go to work on public holidays'.*

The concept *'purchasing intent'* could be operationalized by indicating on a scale of 1 to 5 the likelihood of purchasing the item in the next seven days. We could define *'career success'* either from the perspective of the profession or from the individual's perspective.

The former would define it in terms of the prestige of the profession, the position and status held and by salary. The more individually subjective perspective would focus on definitions that express how one feels about where one aspired to be at this point in time with where one actually is, using questionnaire items such as *'to what extent do you feel you have achieved whatever you had hoped to achieve at this stage of your career?'*

Operational definitions of *'Job Satisfaction'* would involve questionnaire items tapping various dimensions of satisfaction to do with pay, supervision, promotion, etc. Such questions rated on a five-point scale might include:

To what extent would you agree with the following:

- Opportunities for advancement are good here.
- My work gives me a sense of accomplishment.

There is no single acceptable way to operationalize a concept. Suppose we want to measure how depressed an employee is in a work stress study. A variety of options exist:

- We could ask the person if they are depressed (yes or no) or to rate their depression on a 1 to 7 scale.
- We could ask the person's parents or friends if he or she could be characterized as depressed.
- We could have the person evaluated by a clinical psychologist to see if there are clinical signs of depression.
- We could have the person take a test such as the Beck Depression Inventory (a standardized self-report questionnaire designed to measure depression).
- We could ask the person how many serious life events he or she recently experienced that we know often precede depression (such as job loss, death of a spouse, failing an exam).
- We could examine how much of a neurotransmitter called *serotonin* the person has (depressed people tend to have deficiencies in this neurotransmitter).
- We could evaluate glucose metabolism with a technique called Positron Emission Tomography (depressed people often don't metabolize glucose very actively).

The important thing you should realize is that operational definitions turn the vague construct into something that can be measured and quantified.

Choosing an operational definition

As illustrated above, there are often many different ways of operationalizing a construct. But which one is best? There are three broad answers.

1. Use the definition or definitions that work for you, depending on the resources you have or to which you have access. Some operational definitions, such as self-reports, produce data easily. It is fairly easy to get someone to fill out a questionnaire. Other operationalizations might be more difficult. For example, you may not have the money to hire clinical psychologists to evaluate every participant in your depression study, with several hours of clinical interview required per subject. Similarly, access to medical equipment required to measure physiological markers of depression is not easy.
2. Use the operational definition or definitions that will be convincing to those who will be reading and evaluating your research. Use operationalizations that have been employed by other people before, then if someone criticizes your methodology invoke precedent. (Of course, the previous study may have been less than ideal and subject to criticism as well!).
3. Use more than one operationalization. Each operational definition is likely to be incomplete, capturing only one part of the meaning of your construct. By using more than one operational definition, you are better able to cover the richness and complexity of your construct. You will also be able to calculate the *reliability* and *validity* of your measurement, which would indicate that you are obtaining the same result on several different operationalizations of the construct. This would provide converging evidence of your construct. You could also aggregate your measurements on the different operationalizations (adding them up).

Remember that one of the philosophies underlying research is that we can disprove but not prove. Instead, each study gives us one small piece of evidence in a large and complex puzzle. It is now easier to see why this is true. Because there are many ways of operationalizing constructs being studied, it is always possible that any results found may be specific only to that single operationalization, and that had a different operationalization been used, a different result may have been found.

(Burns & Burns, 2008, pp. 74-76)

How to operationalize research questions

The process of operationalization is critical for effective research. Operationalization means specifying a set of operations or behaviours that can be measured, addressed or manipulated. What is required here is translating a very general research aim or purpose into specific, concrete questions to which specific, concrete answers can be given. The process moves from the general to the particular, from the abstract to the concrete. Thus the researcher breaks down each general research purpose or general aim into more specific research purposes and constituent elements, continuing the process until specific, concrete questions have been reached to which specific answers can be provided. Two examples of this are provided below.

Let us imagine that the overall research aim is to ascertain the continuity between primary and secondary education (Morrison 1993: 31-3). This is very general, and needs to be translated into more specific terms. Hence the researcher might deconstruct the term ‘continuity’ into several components, for example experiences, syllabus content, teaching and learning styles, skills, concepts, organizational

arrangements, aims and objectives, ethos, assessment. Given the vast scope of this the decision is taken to focus on continuity of pedagogy. This is then broken down into its component areas:

- the level of continuity of pedagogy
- the nature of continuity of pedagogy
- the degree of success of continuity of pedagogy
- the responsibility for continuity
- record keeping and documentation of continuity
- resources available to support continuity.

The researcher might take this further into investigating: the nature of the continuity (i.e. the provision of information about continuity); the *degree* of continuity (i.e. a measure against a given criterion); the *level of success of the continuity* (i.e. a judgement). An operationalized set of research questions, then, might be as follows:

- How much continuity of pedagogy is occurring across the transition stages in each curriculum area? What kind of evidence is required to answer this question? On what criteria will the level of continuity be decided?
- What pedagogical styles operate in each curriculum area? What are the most frequent and most preferred? What is the balance of pedagogical styles? How is pedagogy influenced by resources? To what extent is continuity planned and recorded? On what criteria will the nature of continuity be decided? What kind of evidence is required to answer this question?
- On what aspects of pedagogy does planning take place? By what criteria will the level of success of continuity be judged? Over how many students, teachers or curriculum areas will the incidence of continuity have to occur for it to be judged successful? What kind of evidence is required to answer this question?
- Is continuity occurring by accident or design? How will the extent of planned and unplanned continuity be gauged? What kind of evidence is required to answer this question?
- Who has responsibility for continuity at the transition points? What is being undertaken by these people?
- How are records kept on continuity in the schools? Who keeps these records? What is recorded? How frequently are the records updated and reviewed? What kind of evidence is required to answer this question?
- What resources are there to support continuity at the point of transition? How adequate are these resources? What kind of evidence is required to answer this question?

It can be seen that these questions, several in number, have moved the research from simply an expression of interest (or a general aim) into a series of issues that lend themselves to being investigated in concrete terms. This is precisely what we mean by *the process of operationalization.* It is now possible to identify not only the specific questions to be posed, but also the instruments that might be needed to acquire data to answer them (e.g., semi-structured interviews, rating scales on questionnaires, or documentary analysis). By this process of operationalization we thus make a general purpose amenable to investigation, e.g. by measurement (Rose and Sullivan, 1993: 6) or some other means. The number of operationalized research questions is large here, and may have to be reduced to maybe four or five at most, in order to render the research manageable.

An alternative way of operationalizing research questions takes the form of hypothesis raising and hypothesis testing. A ‘good’ hypothesis has several features. First, it is clear on whether it is directional or non-directional: a directional hypothesis states the kind or direction of difference or relationship between two conditions or two groups of participants (e.g., students’ performance *increases* when they are intrinsically motivated). A non-directional hypothesis simply predicts that there will be a difference or

relationship between two conditions or two groups of participants (e.g., there is a *difference* in students' performance according to their level of intrinsic motivation), without stating whether the difference, for example, is an increase *or* a decrease). (For statistical purposes, a directional hypothesis requires a one-tailed test whereas a non-directional hypothesis uses a two-tailed test.) Directional hypotheses are often used when past research, predictions, or theory suggest that the findings may go in a particular direction, whereas non-directional hypotheses are used when past research or theory is unclear or contradictory or where prediction is not possible, i.e., where the results are more open-ended.

Second, a 'good' hypothesis is written in a testable form, in a way that makes it clear how the researcher will design an experiment or survey to test the hypothesis, for example, people perform *a mathematics task better when there is silence* in *the room than when there is not.* The concept *of interference* by noise has been operationalized in order to produce a testable hypothesis.

Third, a 'good' hypothesis is written in a form that can yield measurable results. For example, in the hypothesis *people* work better in quiet rather *than* noisy *conditions* it is important to define the operations for 'work better', 'quiet' and 'noisy'. Here 'perform better' might mean 'obtain a higher score on the mathematics test', 'quiet might mean 'silence', and 'noisy' might mean 'having music playing'. Hence the fully operationalized hypothesis might be people obtain *a higher score* on *a mathematics test when tested when there is silence rather than when there is music playing.* One can see here that the *score* is measurable and that there is zero noise, i.e., a measure of the noise level.

In conducting research using hypotheses one has to be prepared to use several hypotheses (Muijs 2004: 16) in order to catch the complexity of the phenomenon being researched, and not least because mediating variables have to be included in the research. For example, the degree of 'willing cooperation' (dependent variable) in an organization's staff is influenced by professional leadership (independent variable) and the personal leadership qualities of the leader (mediating variable: Mastrangelo *et al.* 2004) - which needs to be operationalized more specifically, of course.

There is also the need to consider the null hypothesis and the alternative hypothesis in research that is cast into a hypothesis testing model. The null hypothesis states that, for example, there is no relationship between two variables, or that there has been *no* difference in participants' scores on a pretest and a post-test of history, or that there is *no* difference between males and females in respect *of* their science examination results. The *alternative* hypothesis states, for example: there *is* a correlation between motivation and performance; there *is* a difference between males' and females' scores on science; there *is* a difference between the pretest and post-test scores on history. The alternative hypothesis is often supported when the null hypothesis is 'not supported', i.e., if the null hypothesis is not supported then the alternative hypothesis is. The two kinds of hypothesis are usually written thus:

H_o : the null hypothesis
H_1 : the alternative hypothesis.

(Cohen, Manion & Morrison, 2007, pp. 81-83)

Designing a Research Project Within the Experimental Approach

Finally, let us see how a broad question about language could be turned into the precise form of an experiment. A suitable issue might be the perennial dispute about the relationship between speech and writing. This debate can be carried out in the form of questions.

1. Is speech more important than writing?

The answer "Yes" to this question formed part of the foundation for one school of linguistics, Bloomfieldian structuralism, and for one influential language teaching method, audiolingualism; indeed an affirmative answer is still assumed by most later schools of linguistics and language teaching theories, as shown for example by the assumption common among communicative language teachers that communication should be channelled through speech. What was this affirmative answer based on?

The justification usually presented for the primacy of speech is from history and language development (Bloomfield, 1933); in the prehistory of the human race speech occurs before writing; societies still exist with a spoken but not a written language; in the development of the child, speech invariably precedes writing and is acquired with ease by virtually all children; writing, however, has to be taught and some children acquire it with difficulty. Householder pointed out that this confused two issues - the way language develops and the way it is used by adult speakers (Householder, 1971). Language development is beside the point so far as the relationship of speech to writing in the mature native speaker is concerned, however interesting it may be in its own right; what matters is how they are connected now. On these lines he indeed argued that writing is dominant in adults' usage rather than speech. He was suggesting that the evidence provided an answer to the question "Does speech come before writing in the development of the human race and the human child?" rather than "Is speech more important than writing?" The applicability to second language learning is even more suspect; why should information about children learning their first language or Egyptians inventing hieroglyphics be relevant to people learning a second language? The audiolingualists' agreement with Bloomfield was premature; L2 learning is not directly concerned with either language history or first language acquisition. The relevance of Bloomfield's arguments for second language teaching depends at best on a strong analogy between first language learning and second language learning, which should not be assumed in advance. What is required is evidence to settle a slightly more precise question, namely:

2. Is speech more important than writing in L2 learning?

L2 learning itself has many aspects, each of which might be affected differently by speech and writing; the question needs to specify which aspect we are interested in. One of the crucial axioms of audiolingualism was that a word should always be introduced in speech rather than in writing. Let us assume, then, that vocabulary is the vital aspect involved. While gaining precision, we have inevitably lost generality; it may be that vocabulary learning is a peculiar issue of its own, unlike other aspects of L2 learning. Taking vocabulary acquisition as a particular test case the question can be recast as:

3. Is speech more important than writing in the L2 learning of vocabulary?

There are still uncertainties to be resolved in this version. Most importantly, the question is phrased in terms that apply to all learners. Perhaps, indeed, all people learn second language vocabulary in the same way. Perhaps, on the contrary, there is great individual variation according to some factor in the learner's mental makeup; it might be a matter of cognitive style or motivation; it might be that older learners depend more on visual aids, perhaps. Let us therefore arbitrarily choose to study adults, thus eliminating at least some of the possible variations.

Question 3 also lacks specification of the mother tongue and the second language involved. Quite possibly the relationship between the two languages will have some effect on the learning of vocabulary. In the case of speech and writing, one particular problem is the type of written script; the relative importance of speech and writing may depend upon whether the script is ideographic, like Chinese, or semi-phonetic, like English. A further problem is the question of literacy, regardless of script; a learner who can read may bring certain preconceptions to second language learning, such as, say, the overall importance of the written language. The question lacks details of the mother tongue and target language involved. Let us then decide arbitrarily

to study educated French learners of English. The question now reads:

4. Is speech more important than writing in the learning of English vocabulary by educated French adult learners?

The precise situation of the learners could also affect the result. Learners in classrooms are treated in a bewildering variety of ways, each of them is in some way involved in the question of speech versus writing; learners outside classrooms meet speech and writing in different proportions according to their jobs, to whether they are in their own countries, and so on. Let us decide to concentrate on one particular type of situation, say, technical schools in France. The question now reads:

5. Is speech more important than writing in the learning of English vocabulary by educated French adult learners being taught in technical schools in France?

In terms of an experimental approach the next step is to set up two groups of learners so that the only difference between them is that between reading and writing. To operationalize question 5 we need two situations where only vocabulary learning is involved and where the only difference between them is oral versus written presentation - the controlled situation of Paivio and Begg (op. cit). In this case we might consider a straight experiment of short duration in which two randomly chosen groups at the technical school are taught vocabulary in the different ways and their success assessed. Or we might take two otherwise equivalent classes in the technical school and require one to be taught orally, the other through writing; clearly this already introduces all kinds of complications—will the students in the two groups actually be equivalent? Will the teachers handle both methods equally well? And so on. Even if this ideal cannot be achieved, at least we should be aware of the other factors that might cause the effects we are concerned with. Let us assume then that we set up an experimental situation in which two groups of learners are taught vocabulary in speech and in writing respectively.

6. Is a group of educated French adult learners of English in technical schools in France who are taught orally better at learning vocabulary than an otherwise identical group who are taught through writing?

The next problem to consider is how we can measure whether one group is better than the other. Again, our decision will reflect our model of language and our view of what is important: Range of vocabulary? Ease of use? Speed of reaction? Fluency of expression? And so on. We might collect specimens of authentic data from the learners and subject it to an analysis of the vocabulary. Or we might decide to invent our own test to cover the particular point we are interested in. Or to use an existing test, say, the English Picture Vocabulary Test (EPVT), a standard test of vocabulary development. But the point is to have a yardstick that will impartially measure the relevant aspects of vocabulary in both groups. If we adopt the EPVT, the question now reads:

7. Are the scores on the EPVT of a group of educated French adult learners of English in technical schools in France who are taught orally better than those of an otherwise identical group who are taught through writing?

Finally, we still need to quantify the word "better": how much do one group's scores need to differ from the others to be called "better"? Partly, this is a matter of statistics and levels of significance; we need to be certain that the results are not sheer chance and this can be done through various mathematical tools. Our design and system of measuring must then permit the use of such tools. Partly, it is a matter of commonsense and interpretation; do we feel that the difference we find is actually important to language learning, whatever the statistics may say? Hence the ultimate form of our question is:

8. Are the scores on the EPVT of a group of educated French adult learners of English in technical schools in France who are taught orally significantly better than those of an otherwise identical group who are taught

through writing?

This quest has tried to narrow down the research question so that it can receive a reliable answer. Instead of the overall, untestable proposition about speech and writing there is now a precise, answerable question. The technique of breaking a problem down into a single testable issue has a long history in science as the "method of detail" (Pacey, 1974). Galileo, for instance, tackled the problem of the strength of ships' hulls by reducing it to the strength of a single beam. The decomposition of a problem into elements is also often discussed as one of the benefits of teaching computing to children (O'Shea and Self, 1983). It is not then an approach unique to second language learning research.

(Cook, 1986, pp. 16-19)

Operationalization

Most social science research involves making observations that we presume tap concepts. If we were conducting a study on the effect of marital breakdown on the wellbeing of children we would need first to work out what is meant by marriage breakdown, wellbeing and children. This involves defining these concepts, which in turn requires developing a *nominal definition* and an *operational definition* of each concept.

Concepts are, by their nature, not directly observable. We cannot see social class, marital happiness, intelligence etc. To use concepts in research we need to translate concepts into something observable something we can measure. This involves defining and clarifying abstract concepts and developing indicators of them. This process of clarifying abstract concepts and translating them into specific, observable measures is called operationalization and involves *descending the ladder of abstraction.*

(de Vaus, 2001, p. 24)

3.3.1 Research topic, research purpose, and research questions

Every investigation has a starting point and unless we adopt someone else's (for example, the supervisor's) design idea or join an ongoing investigation as a co-researcher, this starting point is a broad 'research topic'. Initially, this is only a general and rather vague area of research interest (for example, a problem or an intriguing phenomenon that we have noticed) and in order to be able to specifically address it we need to develop it into a 'research purpose'. The research purpose is a relatively short statement (not more than a paragraph) that describes the objective of the planned study, explaining why the investigation is undertaken and what its potential significance is.

The next step towards narrowing down and 'operationalizing' the research purpose is to formulate specific 'research questions'. Their role is to translate the research purpose into specific questions that the planned study will aim to answer. Being able to produce good research questions indicates that the researcher is ready to embark on the actual investigation (whereas vague or missing research questions usually send alarm signals to a supervisor or referee), and good research questions also provide orientation to the research methodology that can best achieve the research purpose. We should note here that, as Johnson and Christensen (2004) point out, because in some investigations the research question would be, to a great extent, a restatement of the purpose of the study, some scholars actually omit it from the final research report.

What is a good research question like? As we will see in the next section, there is a difference between effective research questions in QUAL and QUAN studies, but in general, good research questions need to address *interesting* issues. When we have thought of some possible questions, we have to ask ourselves: are they worth asking and, more importantly, answering? I believe that one thing we must try and avoid at all cost is to 'run into a "so what" response to our research' (Mackey and Gass 2005: 17). Therefore, I am in agreement with Gall *et al.'s* (2007: 41) conclusion that 'The imagination and insight that goes into defining the research problem usually determines the ultimate value of a research study more than any other factor'.

Where can we receive the inspiration for good research topics and questions? In my experience, most research topics originate from a combination of reading the literature and one's personal history. From time to time something we read 'rings a bell' and we realize that the topic could be further pursued on the basis of relevant experiences we had in the past. In addition, the Conclusion section of most research articles tends to contain suggestions for further research, and discussions with friends, colleagues and students can also be very helpful in drawing our attention to potentially fruitful issues. If you decide to keep a research journal, the reflections recorded in the journal may contribute to refining the research questions of an ongoing investigation as well as generating ideas for future research.

Sometimes, when we have a hunch or a learned guess about the possible results of our planned investigation, we can narrow down the research purpose into actual 'research hypotheses' instead of, or besides, questions. These hypotheses are statements that formulate specific predictions about the outcomes and the empirical results will either confirm or refute these. If generating research hypotheses is possible in a study, they are very welcome because they are fully in line with the principles of the 'scientific method'.

Nunan (1992) makes a useful recommendation about how to operationalize an investigation. He has found that asking postgraduate students to create a 'research outline' can greatly facilitate the planning and fine-tuning of a project. Such an outline consists of a series of headings, and students are to add a short statement after each; the headings can be selected in a flexible way, but they would normally include the general research area, research purpose, research question/hypothesis, as well as data collection method, type of data, participants, type of data analysis, and resources required.

(Dörnyei, 2007, ch. 3.3.1)

Validity/Trustworthiness/Reliability

With all research, we need to be able to be confident in our results; our results need to be trustworthy and valid. Validity encompasses many of the concepts already discussed in this chapter. In the following section, we refer to a concept that is often discussed together with validity and that is reliability. Validity refers to the correctness and appropriateness of the interpretations that a researcher makes of his/her study. Reliability refers to score consistency across administrations of one's instrument. For both concepts, accurate and appropriate instruments are at the core. Thus, if our hypothesis involves learning, we need to have an accurate and appropriate instrument to measure learning. For example, if we are looking at knowledge representation, we need to have a measure that appropriately reflects that and not something that just measures an ability to use the language. In many instances (see discussions in Gass and Mackey 2007; Gass and Selinker 2008), part of the theoretical discussions in the literature involve how best to represent particular constructs.

Often variables cannot be measured directly. In these instances, we come up with a working definition that allows us to identify the variable in question with something that is understandable and measurable. This is known as an operationalization. Thus, in our example presented earlier of focused attention, we cannot

directly measure this construct but we can come up with a reasonable surrogate (e.g., colouring or highlighting in some way). Once we have operationalized a variable, we can more easily work with it.

The last thing that a researcher wants is to spend time, effort and money on a project and then realize that the study itself did not reflect what we had thought it would and might only apply to the population of the study and not to the broader community at large. Validity comes in many different colours; in this section we discuss the most common types of validity: content, face, construct, criterion-related and predictive validity. Following that brief introduction, we turn to a discussion of internal and external validity.

Content validity refers to the representativeness of our measurement regarding the phenomena that we want information about. If, for example, we want information about noun-adjective agreement in Italian, we cannot generalize these findings to say that we have fully investigated all types of agreement (article-noun, singular-plural, regular-irregular, nouns that end in —*a-o* and those that are not morphologically marked as masculine/feminine). In other words, if we want to claim that we have investigated agreement, we need to ensure that our instruments include a representative range of what constitutes agreement.

Face validity is closely related to the notion of content validity and takes us into the realm of the consumers of research. Is our instrument readily recognizable as measuring what we claim it measures? For example, the construct of intelligence can be measured in various ways, but there are certain instruments that are well-accepted as measuring this construct, even if it is a somewhat elusive construct. Thus, face validity refers to the familiarity of our instrument and how easy it is to convince others that there is content validity to it. If a school district wants to measure intelligence with a newly developed instrument, there may be a perception by the community (e.g., parents) that this instrument is not valid (unless of course their child receives a high score!). If the participants do not perceive a connection between the research activities and other educational or second language activities, they may be less likely to take the experiment seriously.

Construct validity refers to the extent to which the research adequately captures the concept in question. In second language and applied linguistic research, construct validity is of great concern because a great deal of what we investigate is not easily quantifiable and not directly measurable. Some variables, such as height, weight, shoe size are easily measurable and there is little controversy over what they reflect. Thus, a height of 5’5" is clear to everyone who uses this scale, but in second language research, we are dealing with such constructs as proficiency. What precisely does this mean? How can we measure it so that we can compare individuals on a common scale? Because these constructs are not directly measurable, their validity can be called into question. One way to enhance construct validity is to have multiple measures. Thus, if we were to measure proficiency, we might have measures that reflect oral use, written use, extent of vocabulary knowledge and so forth. If we were to use these measures as an aggregate, we could have greater confidence in our ability to differentiate individuals along a scale of proficiency.

Criterion-related validity refers to the relationship that a given measure has with some other well-established measure. For example, if a researcher develops an overall measure of language proficiency, it will have criterion-related validity if it measures language students in much the same way as another well-established test. To be more specific, if we are doing a study using first, second and third year English learners of Spanish and we developed a test that measures oral proficiency, criterion-related validity would be increased if we could show that on our test, third-year students did better than second-year students who did better than first-year students. Our test, then, would correspond to some other reasonably accepted measure, that of class placement.

Predictive validity deals with how well the measure we are using predicts performance on some other measure. In other words, if we have a test that measures working memory capacity, it has predictive validity if it predicts performance in class performance in a language class.

In addition to these five types of validity, there are two other types of validity that are noteworthy: internal validity and external validity. Each of these is important when conducting experimental research.

To what extent are the results of a study truly reflective of what we believe they reflect? This is known as internal validity. In other words, are our dependent and independent variables related in the way we think they are? A researcher must control for (i.e., rule out) all other possible factors that could potentially account for the results. This was discussed in the example above in relation to colour blindness. In that study, had we not controlled for colour-blindness, we would be left with the unfortunate conclusion that the study had little internal validity. Before conducting any research study, we need to think carefully through the design to ensure that we eliminate or minimize threats to internal validity (see Mackey and Gass 2005 for a more thorough treatment of this topic and for ways to minimize threats to internal validity).

External validity refers to the potential generalizability[2] of a study. We can make conclusions about the behaviour of the participants in a study, but this is not particularly interesting unless the results have broader implications and are relevant to a wider range of language learners and language learning contexts.

Thus, if we conduct a study with English-speaking learners of Italian studying at University ABC, we are interested not just in those specific learners, but also the extent to which the results are applicable to learners of other languages possibly in different contexts. This is the case because we are interested in general principles of learning and not just a particular group of learners.

External validity can be increased with appropriate sampling procedures, as mentioned above. In particular, it is important that our sample be selected randomly which, in essence, means that each member of the population to be studied has an equal and independent chance of being selected. This is the ideal situation, but one which, in reality, is not always practical. Rather, in second language research, non-random sampling is frequently used. Researchers often seek volunteers to participate in a study, as is required by most university ethics review boards. Even when intact classes are used, students can opt out of participation according to university ethical requirements. Sufficiently large sample sizes are always a goal as a way to increase the likelihood of true differences between groups (e.g., experimental and control). Small sample sizes leave the researcher with the uncertainty of understanding the results. Are the differences between groups true or just coincidental? Many statistical tests help researchers avoid drawing unwarranted conclusions.

Because true random sampling is not always likely in second language research, it is important for researchers to fully and accurately describe the population studied as well as provide details about the materials, methods and procedures. In this way, a particular study can be replicated by others which, in a way, broadens the population base of the original study (see Polio and Gass 1997 for a fuller description of replication and reporting). Mackey and Gass (2005) provide additional discussion on issues of external validity and outline ways of minimizing threats to external validity.

Reliability refers to consistency and is a way of ensuring that our constructs are being measured appropriately. In applied linguistics research, it is frequently used when raters are making judgements about data. This is referred to as interrater reliability (when more than one rater is involved) and intrarater reliability (when only one researcher's evaluations are used). In the former instance (e.g., judging oral speech samples on a scale of 1-10), consistency across raters indicates that raters are measuring the same construct in the same way. In the latter case, one might rate the same speech

sample at two different points in time to ensure consistency.

Both validity and reliability are ways of ensuring quality in research. As noted, experimental research is a way of finding answers to questions in a disciplined way. These results may have far-reaching impact (including decisions relating to educational practices), and it is incumbent on the research community to ensure that research (experimental and other) is carried out in as careful a way as possible, ensuring quality at each step of the way.

(Gass, 2010, pp. 12-15)

WHAT IS "A SYSTEMATIC WAY"?

A systematic way may once again remind you of laboratory experiments with Marie Curie carefully recording weights under extremely controlled conditions in her laboratory. To be sure, laboratory experimentation is one of many systematic ways to conduct research, but it is by no means the only way. By a systematic way we simply mean that the researcher should follow established principles. This is a matter of avoiding ad hoc solutions during the investigation. By clearly outlining your procedure and maintaining consistency, you can reduce the effect of your personal preferences as well as other extraneous factors which might influence the outcome of the research.

It is not the case that only laboratory experiments are systematic. The systematicity of procedures can and must be maintained in any type of research. Case studies, longitudinal and cross-sectional studies, critical library research also have principles to follow. In short, no matter what kind of research you arc involved in, certain conditions must be present in order for your results to be valid and useful.

In addition, the research method should be the most efficient way of answering your question. It’s no use to formulate a question and then collect data in a way that will not allow you to answer the question. It is important to select the approach that is most appropriate to answering the research question. In the relative clause example, the data could come from case studies, questionnaires, student compositions, tests, or experiments. Each approach would have strengths and weaknesses that you should consider before you make a final choice of procedure.

No matter which approach you use, if the hypothesis is that there is no relationship between learners’ first language and acquisition of relative clauses (i.e., similarity/difference of the two languages has no effect on learning), then it is crucial to have a clear definition similarity no matter what research method is used. An operational definition already justified in previous research may be used. If not, then the researcher will have to propose and justify one. Otherwise, inconsistent and biased decisions are likely to be made during the research no matter what approach is decided on. A decision must also be made regarding the kinds of relative clauses to be included in the study. They must be clearly identified. And finally, regardless of what approach you use, you must decide how strong a relationship (whether negative or positive) you will have to find before you can say that the null hypothesis is rejected. We will talk about how we make this decision later. The point is that you must be systematic or there will be so many weaknesses in the research that later you won’t know how to go about plugging up the holes. As a result, your work won’t be worth much to you or to anyone else.

The researcher should consider which approach will be most efficient in giving answers to the research questions. For example, if you decided to use a case study approach to investigate relative clauses, you would most likely be disappointed. If your data were not already collected and transcribed, you would first have to learn the conventions for collecting data and transcribing them. This would involve many, many hours of your

time. Once transcribed, you could not expect to find many relative clauses in the data. You could tally the numbers of relative clauses over several months but it is unlikely that all the clause types you had so carefully defined ahead of time would actually appear in the data. Given the paucity of relative clauses in the data, you would probably wonder why and turn the focus of the study toward understanding the function of relative clauses in spoken discourse. This would mean revision of your operational definition of similarity between first and second languages. Now you would need to know what constitutes similarity in function as well as in form of relative clauses. In short, you might start with a perfectly good research question, state your hypotheses, and then select a method that won't let you answer the question as it was originally formulated. The strength of the approach would be that you have real production data. The weaknesses would be that you have so very little data from only a few subjects that you cannot really answer the question you have asked.

If you decided to use a test in order to get the information you need to answer the question, you would need the same specific information about types of relative clauses to include and the operational definition of what constitutes similarity/difference for the first and target language forms. You would also need to have students from a variety of first language groups, and you would need to specify the level of proficiency already attained by the students. The strength of choosing this method is that you will get lots of relative clause data and, if you construct the test carefully, examples of each clause type you wish to include. It will be an efficient approach, for you could probably collect the data at one sitting. Depending on the form of the test, you will probably be able to say something about whether or not students from these first language backgrounds are able to recognize correct forms. It cannot tell you whether students can accurately produce relative clauses or whether they use them appropriately in the real world of discourse. The approach will be efficient, but it will not allow you to answer the research question unless you change the hypothesis to claims about recognition as equivalent to learning.

If you chose an error analysis approach, you might collect student compositions and hope (as in the collection of oral data in case studies) that you will find enough instances of relative clause use (and of each clause type) to provide you with data that speak to the question. The strength of the method is that you will have natural, production data and that written discourse makes heavier use of relative clauses than oral, spoken discourse. If you are lucky, the composition topic might be one that would require a great deal of identification and definition of nouns (and thus many relative clauses). That this might be an effective compromise between the weaknesses of the case study approach and the test approach is obvious from the many studies which have looked at relative clauses using this approach.

There are, clearly, many different approaches to take in answering any research question. None of these methods should be preferred over the others just because one method is easier than another, because your friend volunteers her class for an hour, or because you just met a speaker of some Nilotic language who is about to utter his first relative clause. Selecting one particular approach should depend on the nature of the research question and the hypotheses that you have made. And, of course, a multimethod approach is the best of all for it will allow you to feel more comfortable about any claims you want to make. It's always comforting to find that you get the same results when you use another approach as well.

Each of the approaches has conventions which should be followed. If you disregard them, you will soon find out why the conventions were established in the first place. The conventions are the result of many other researchers trying to avoid the mistakes which invalidate results. If you follow the conventions, you will avoid mistakes others have made. There will still be plenty of opportunity for you to make your own! If you select the most appropriate research method and follow the conventions which make it systematic, you should be able to make valid statements about the results of your study when you finish.

Let's forget about relative clauses for the moment and turn to a different hypothetical question. Suppose you were interested in answering the question of whether the reading ability of your ESL students improved over

a term of instruction in which you used a marvellous new reading method which you had invented (and wanted to sell to some publisher). Fortunately, the students had been given a reading test at the beginning of the term, and at the end of the term, you gave another test. Comparing the scores of the students on the two tests, you conclude that their reading ability, has significantly improved. But is your conclusion a valid one? Could you conclude that your new method improved the students' reading ability? Could you be sure it was the instruction and not other factors that promoted improvement? Could you make statements about the relationship between the degree of improvement and the amount of instruction? Could you convince a publisher who is knowledgeable about design and statistics?

These questions are concerned with the third key term in the definition of research, finding answers. When we investigate a research question, the findings provide potential answers to the question. However, no meaningful interpretation of the results can be made unless the data-gathering procedures have been done with care. If the procedures are flawed. then neither the results nor the interpretation of the results can be valid. Validity of research is a crucial concept.

(Hatch & Farhady, 1982, pp. 4-7)

Scope of Research

Aside from personal interest, research questions need to have other characteristics. They should be able to generate new information or confirm old information in new ways. To be sure that this is the case, a review of the literature on the research topic must be done. Imagine that you are interested in second language proficiency but instead of looking at students' use of grammatical structures, you want to investigate how well they can perform basic "speech acts." Obviously this is not a topic that one would select without some acquaintance with speech act theory. You also would not select the topic if you thought that it had already been sufficiently addressed.

Still, the first thing to do is undertake a review of previous speech act research to learn exactly what has been done in this area with second language learners. If you went to your university library, you might be able to get a computer search for this topic. Such searches, like ordinary library searches, begin by looking at "key words" and "key authors." Key words and key authors for speech acts might include terms such as *directive, assertive, commissive* or such authors as *Austin, Searle, Gumperz.* Think for a moment about how broad a key word like "speech act" might be in such a search. While it is not as broad a key word as, say, *linguistics,* the number of articles and books generated by a search with this key word would be very large (and very broad).

A search using *bilingual* as a key word would also generate a huge number of items. Many of these would not be useful for, say, a language policy study of bilingualism in Peru. The question is how to narrow the scope of the search and at the same time find all the relevant entries.

Hopefully, this question illuminates one of the first problems regarding the definition of research questions - the questions are stated too broadly. To say we want to know more about how well second language learners carry out speech acts is a little like saying we want to know how; learners use language. Precisely *which* speech acts to we want to investigate? What types of second language learners - beginners, advanced – are we talking about? Are the learners adult Korean immigrants in Los Angeles or Japanese high school students in Kyoto? In what kinds of situations should the speech events be investigated? Is the research meant to support test development? materials development? theory development? Where above we called for narrowing via key words, now we can narrow further via key sentences.

Using these questions as a guide, we can redefine the research question by narrowing the scope. For example, the scope could be narrowed from:

> *Investigate how well second language learners perform speech acts.*

to:

> *Investigate Korean ESL students' ability to recognize complaint behavior appropriate in an academic university setting.*

Here "performance" has been narrowed from total performance to recognition (one portion of the total performance skill). "Second language learners" is narrowed to "Korean ESL students," and "speech acts" has been narrowed to one speech act subcategory "complaints." The events in which the subcategory might occur have been narrowed to those relevant to the university setting. There are, of course, many other ways in which the question might be narrowed.

In narrowing the scope of the research, we may lose interest in the topic because it no longer addresses the larger question. An appropriate balance needs to be struck between scope and interest. It is possible to maintain the original research interest by carrying out a number of studies with limited scope. Together these studies would address the broader, general area of interest.

A review of previous research will help us to define the scope of research in another way. We've already noted that the scope of the research must be realistic. But, even a fairly narrow question may need to be more carefully defined. Previous researchers may already have done this. For example, many teachers are concerned with student motivation. Motivation, like bilingualism, is a very broad concept. Previous researchers have, however, grappled with this problem and have subcategorized the concept into types of motivation--for example, *intrinsic* and *extrinsic* motivation or *instrumental* and *integrative* motivation. In narrowing or subcategorizing the concept, operational definitions must be given to show the scope of the subcategory.

Sometimes well-established *operational definitions* exist for terms that are crucial to your research. Such a definition gives a "tried and true" definition and an accepted method for observing or assessing it. However, sometimes there are no such accepted definitions or no agreement as to what the terms mean. There are, for example, many abstract theoretical concepts that have been "constructed" in our field. These *constructs* are shown in abstract terms such as *acquisition, motivation, need, achievement, monitoring, compound bilingualism.* We may share a basic understanding of such theoretical concepts, but even these *theoretical definitions* are difficult to formulate. For example, precisely *how* would you define *bilingual?* A commonly-shared definition of *bilingual* is "speaking two languages." We all know that the term may be applied to people who are at all points of fluency in the two languages (even to absolute beginners of a second language). To use such a term in research would be almost meaningless. A more precise definition is given, for example, *Arabic-English bilinguals who scored a 3+ or higher on the FSI inventory participated in this study* or *Children who had participated in the Arabic immersion program in Cleveland schools in grades K-3 constitute the bilingual group in this study.*

When broad terms for constructs are used in research questions, we cannot rely on a *theoretical definition* even if one is readily available. Terms must be "operationally" defined. An *operational definition* is a clear statement of how you judged or identified a term in your research. This is important for three reasons. First, you will need to be absolutely consistent throughout the research process in your definition. Second, it is important for consumers of your research so that they do not misinterpret your findings. Third, it is important to the research community that your

study be replicable. Different results might be obtained by other researchers if they carry out a similar project and use a different definition of *bilingual.*

Good operational definitions can often be drawn from the existing literature. Sometimes, however, research is difficult to carry out because operational definitions cannot be found that will satisfy the researcher. Sometimes no operational definitions exist in the literature and the researcher must define terms. We know very little, for example, about how language is represented in the brain. Yet, many models of language acquisition talk about "acquisition devices," "filters," "parameters," "L1 → L2 transfer" as *internal* mechanisms. It is, of course, possible to create and define an operational definition for these terms for an individual project. A clear definition would be crucial to the research. (In some cases, we develop an operational definition for such concepts but then find ourselves questioning the "reality" of the concepts themselves. The attempt to establish concepts is an important area of research.

(Hatch & Lazaraton, 1991, pp. 14-17)

Operationalism

The multiple features of any construct and the lack of clear prototypical features of many constructs create difficulty for researchers when they try to represent constructs of interest in their research studies. What is the researcher to do? The researcher must make use of the available knowledge and measures of the construct he or she is investigating and identify the specific way in which a construct will be represented in the study he or she is conducting. This is where the concept of operationalism enters, and it is an important communication tool for researchers. Operationalism means that terms or constructs are represented by a specific set of steps or operations. For example, if stress is measured by the Stress in Teaching Questionnaire, then the construct of stress is represented by scores on this particular questionnaire (Manthei & Gilmore, 1996).

The researcher should select the best measure available for his or construct of interest. Use of a particular measure should be justified (i.e., reasons given) in the writing of the research report. For example, Butler and Neuman's (1995) representation of help-seeking behaviors in second- and sixth-grade children was whether the children asked the experimenter for assistance in solving puzzles. The researchers, operationalization (i.e., their construct representation) of help seeking left little room for interpretation or question as to the way in which they had conceptualized and interpreted help seeking. Note, however, that this is not the only way in which a person can seek help. Help can also be obtained by asking peers for help or going to the library and looking for reference materials that would provide assistance. The important point to remember is that specification of a set of operations is required for accuracy in communication. This is the beauty of operationalizations. They specify, in a concrete and specific way, how a construct is represented or measured in a given study. This degree of specificity permits an exact communication of the construct and allows anyone else to repeat the steps and represent the construct in the same way.

Although the operationalization of constructs is necessary for communication of the way in which a construct is represented in a given research study, seldom, if ever, does a given operationalization completely represent the construct being investigated. Consider, for example, the study by Manthei and Gilmore (1996) in which they operationally represented teacher stress as the response teachers provided on the Stress in Teaching Questionnaire. Although this questionnaire probably does measure some component of teacher stress, it would be foolish to assume that this single measure provides a completely accurate representation of the construct of teacher stress. Rather, stress of any type probably includes physiological reactions such as

changes in heart rate and blood pressure as well as behavioral changes such as being less tolerant of the students, both of which are probably not adequately assessed by a questionnaire.

Additionally, Donald Campbell (1988) has pointed out that every observation is affected by factors that bear no relation to the construct that is being measured. For example, the Stress in Teaching Questionnaire does, in part, measure teacher stress. However, responses to this questionnaire are also a function of events that are irrelevant to the stress that occurs as a result of teaching, such as the type of questions asked, the interpretation of the questions by the teacher completing the questionnaire, the tolerance a teacher has for stress, and stress factors influencing the teacher that are not related to the profession of teaching. Therefore, measurement always involves the presence of some measurement error.

The important point to remember is that there are many different ways of operationally representing a construct and that each operationalization represents only a portion of the construct. Furthermore, some measurement error is always present. The most accurate representation of a construct involves measuring it in several different ways. For example, teacher stress could be measured by a questionnaire, by the teacher's reaction to students, and by having others rate or identify factors influencing teacher stress. As more and more measures of the same construct are included, the probability of obtaining a more complete and accurate representation of the construct increases. The use of multiple measures of a construct is called multiple operationalism, which is the recommended approach to use in research studies (Campbell, 1988).

We also must point out that it is not sufficient to specify a set of operations as a representation of a construct and then assume that this is a valid measure of the intended construct or even some component of the construct of interest. Simply describing the operationalization of a construct does not guarantee success in representing the construct. All the operations do is state in specific and concrete terms the way in which the construct is being represented in a particular study. To drive this point home with a ridiculous example, assume that you want to investigate the effect of intelligence on learning. In this study, you operationalize intelligence as a person's income on the assumption that more intelligent people make more money. This is obviously a poor representation of the construct of intelligence. However, it is operationalized in specific and concrete terms, and it does communicate how intelligence was represented in the study. A key point is that researchers must provide sufficient evidence that the operationalization of a construct used in a research study adequately represents the construct. As a "critical" consumer of research, it will be your job to check for adequate operationalizations of constructs in the reports that you read and use. You will find his information in the Method section of the report.

Before moving on, we need to mention quickly a construct problem that sometimes occurs in experimental research conducted in the field over a period of time. Because the researcher cannot control who interacts with whom in a field study, treatment diffusion can occur if participants in the treatment group interact or share resources with participants in the control group. If this happens, what is thought to represent the control condition is no longer correct because it has become contaminated with something from the treatment condition. The point is that it can be difficult to keep the levels of your independent variable (e.g., treatment and control) truly separate in a field experiment, and this can be difficult to detect.

(Johnson & Christensen, 2012, pp. 262-264)

Operationalization

The way in which a larger **construct** that is the subject of investigation in a research study is measured. For example, there are many psychological constructs, such as anxiety, motivation, and personality type. However,

it is not possible to measure these constructs directly or exhaustively; rather, a researcher must decide what measures will best represent the construct. Thus, for anxiety, researchers might decide to use a **survey to** operationalize the construct, or they might use a set of behaviors exhibited by learners in a specific context. These measures would constitute the way in which anxiety is operationalized in the study Another common example of operationalization in SLA is the measurement or 12 proficiency: again, It is not possible to examine the level of language ability Inside a learner's head, so proficiency must be represented in specific, measurable ways. For L2 proficiency, researchers often take an institutional placement level as an operationalization of proficiency. For example, in their eye tracking and grammaticality judgment tests, Godfroid et al. (2115) operationalised proficiency by reporting the placement level of their participants (Levels 3, 4 & 5) at the English language center at which they were studying. Researchers might also use tests targeting specific linguistic areas, grammar or vocabulary, or tests of a more general nature, such as TOEFL or IELTS.

It is important to recognize that the way in which a construct is operationalized affects the interpretation of the data. For example, researchers might come to different conclusions if they use different measures of L2 proficiency. Although some methods of operationalization may be better than others because they have higher **construct validity,** there is generally no perfect method of operationalization. Every method will have some shortcoming, such as measuring only part of the construct, Thus, researchers should be clear on how they are operationalizing the construct as well as the limitations inherent to their approach. To overcome such limitations, researchers often use more than one method of operationalization to triangulate and reach firmer conclusions about their data,

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (201 S). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition,* 37, 269-297.

(Loewen & Plonsky, 2016, pp.132-133)

5.4 OPERATIONALIZATION

In many instances in second language research, it is difficult to measure variables directly, so researchers provide working definitions of variables, known as *operationalizations.* An operational definition allows researchers to operate, or work, with the variables. Operationalizations allow measurement. To return to the example above, we said that we need to operationalize *amount of input,* because this term, as stated, is vague. Although it might be difficult to come up with a uniform concept of amount of input, it is possible to think of examples in which groups vary along some parameter that seems close to the amount of input. For example, classroom learners could be classified based on how many years of exposure they have had to the target language. So "amount of input" could be operationalized as years of exposure in this case. In a natural setting, where exposure comes from everyday living, the operationalization of "amount of input" could be the number of years spent in the target language environment. Once a variable has been operationalized in a manner such as this, it is possible to use it in measurements.

(Mackey & Gass< 2022, p. 209)

In addition to internal/external reliability/validity, we need to operationalise the constructs we are working with. In other words, we need to describe their characteristics in a way which would enable an outsider to identify them if they came across them. If researchers fail to provide 'up front' definitions, then we need to read between the lines. For example, if a study investigates 'listening comprehension', and the dependent variable is a written cloze test, then the default definition of 'listening comprehension' is 'the ability to complete a written doze passage'. If we were to find such a definition

unacceptable, we would be questioning the *construct validity* of the study. Construct validity has to do with the question: Is the study actually investigating what it is supposed to be investigating?

In developing a research problem or question, Wiersma (1986) suggests that the problem be broadly stated in the first instance and then progressively refined and restricted through a review of the literature. He provides the following examples of how an initial broad area can be reformulated, either as a problem statement or question.

Original	Achievement and teaching techniques
Restatement	A study of the effects of three teaching techniques on science achievement of junior high school students
Question	Do three different teaching techniques have differing effects on science achievement scores of junior high school students?
Original	Bilingual education
Restatement	A study of the nature and characteristics of bilingual education in the elementary schools of City A
Question	What are the nature and characteristics of bilingual education as it is implemented in the elementary schools of City A?
Original	The role of the guidance counselor
Restatement	A survey of the practices of the guidance counselors in the high schools of City B
Questions	What proportion of guidance counselors' working day is taken up with nonguidance activities? What are the major strengths of guidance counselors' practices as perceived by the students? What are the major weaknesses of guidance counselors' practices as perceived by the students? What practices are perceived by guidance counselors as most effective in advising students about college selection?

(Wiersma 1986: 31-32)

(Nunan, 1992, pp. 214-216)

2.3. Definitions, concepts and operationalization

Some readers may argue that I pay undue interest to something straightforward. After all, most of us are able to measure someone's age, gender or weight. However, there are cases where the process of allocating a variable value to a case is not that easy. A common problem, particularly for researchers new to a field, is that it is not always clear what is being measured, let alone how. A typical example for this is the concept of 'motivation'. Over the years, I have met countless students, both undergraduate and postgraduate, who were all interested in how motivation influences second-language learners' progress. The problem usually starts when I ask the question, 'so how are you going to measure it?' We all know that motivation plays a crucial part in second-language acquisition/learning; however, in order to come to a sound conclusion along the lines of 'the higher learners' motivation, the better the outcome' we need to measure the different levels of motivation. But before we can even measure it, we have to clearly define what motivation actually is. A review of literature on the topic shows that motivation is a complex system of several interrelated factors, of which Stern (1983) identifies social attitudes, values and indeed other 'learner factors'. More recently, motivation is often linked to the 'learner's self' (see Dornyei and Ushioda 2009 for an overview). As such, motivation is a concept rather than a variable; in fact, motivation consists of several variables, each of which

has to be measured separately. In other words, we cannot reliably measure motivation as such, let alone by methods such as asking learners 'how motivated are you'. When constructing concepts, one has to ensure that they are epistemologically sound - put more bluntly, they have to make sense. For example, a concept can only be described by definitions and terms that are already known and measurable. In the above example, we need several already known and measurable learner factors in order to define our concept 'motivation'.

A different example is the concept of 'cohesion': 'Cohesion can be defined as a set of resources for constructing relations in discourse' (Martin 2001: 35). The clue is in the definition: as it is a 'set of resources', it cannot be measured directly, but we can approach it by looking at the different components that make a text cohesive.

Once we have defined the concept we want to investigate, we need to establish a set of operation definitions; we have to operationalize our concept. In other words, we need to decide how we measure the variable or variables describing our concept. To come back to our example, we would develop a set of questions (or variables) which refer to the different aspects of motivation, such as attitudes to the language, the learning process, the environment and so forth.

We shall look at a real example to illustrate the idea of concepts and their measurability. In the late 1970s, Giles et at. (1977) developed the concept of 'Ethnolinguistic Vitality' in order to provide a framework that could account for what 'makes a group likely to behave as a distinctive and active collective entity in intergroup relations' (308). It does not take too long to understand that 'group behaviour' is not something that can be straightforwardly measured, such as a person's age or weight, but is a construct that comprises various aspects and issues. Giles et al. identified three categories of variables, each to be measured independently and, more or less, objectively:

a. The group's social status, based on objective and objectively measurable variables such as income (how much money do individual group members earn), economic activity (unemployment rates), employment patterns (which jobs do they have), but also including more difficult approaches to status, such as perception of the group or the group's language (e.g. whether the group's language is an official language, such as Catalan in Spain).
b. Demographic factors, taking into account the absolute and relative size of the group as well as its density (i.e., how many group members live in a clearly defined geographic space). Again, these variables are comparatively easy to measure.
c. Institutional support, measuring to what extent the group receives formal and informal support through official and/or community-intern institutions (government, schools, churches, local authorities etc). While this last category is a rather abstract concept in itself (for details, see original work), it ultimately provides us with a quantifiable result.

Giles et al. argue that the higher a group fares in each of the categories, the higher its Ethnolinguistic Vitality is. So, the results (or scores) obtained in the individual categories allow us to fairly objectively quantify a group's vitality. In other words, we can put a comparatively reliable number to describe a rather abstract concept.

It is important that, for every quantitative study, we spend some time thinking thoroughly about our variables, how to measure them, and how to operationalize this measurement. And as such, it is paramount that we have a good grasp as to what our variables could potentially look like, before we develop a method to capture as much of the variation in this variable as possible. Phonology is a good linguistic example to illustrate this. Traditionally, sociolinguists and dialectologists have focused on variation in pronunciation, and have tried to explain this variation along various dimensions such as social categories or geographic areas.

Sebba (1993: 23) in his study on 'London Jamaican' discusses some of the phonological differences between Jamaican Creole and London English, and identifies for one particular speaker in his sample two distinct realizations for the pronunciation of the /u/ sound in 'who' and the initial consonant in 'the': [hu:] versus [hʊ:], and [di] versus [ðə]. If, like Sebba, we consider this phonological difference to be important, we must use a measure that accounts for this. In other words, we need an operation definition that allows us - in practice - to clearly classify a particular vowel as [u:]/[ʊ:] and a particular consonant as [d]/[ð]. At the same time, if such differences are not of interest for our study, we do not need to account for them. The trick is to develop a methodology which allows us to measure as accurately as required but does not make things unnecessarily complex.

A word of warning: some concepts, particularly abstract ones and those that comprise several variables and sub-concepts, are notoriously difficult to measure. Motivation, as outlined above, is one such concept, as are attitudes. A quantitative analysis of these takes time, effort and a lot of trial and error. Possibly more than for any other study, the thorough evaluation of previous work and in particular the methodologies used (including all the difficulties and pitfalls!) is vital. It is usually a good although resource-and time-consuming idea to support quantitative data on these issues with qualitative findings from interviews or focus groups. Often, the combination of qualitative and quantitative results provide a far better and reliable insight than one approach on its own.

(Rasinger, 2008, pp. 20-22)

Writing a research aim and set of research objectives

As well as your research question, you may also be required to formulate a research aim. A research aim is a brief statement of the purpose of the research project. It is often written as a sentence stating what you intend to achieve through your research. To illustrate this, the examples of research questions in Table 2.2 have been matched to their research aims in Table 2.3. You will see the close relationship between these - one stated as a question, the other as an aim.

Table 2.2 Examples of research ideas and resulting general focus research questions

Research idea	**Research question**
Media campaign following product recalls	How effective is a media campaign designed to increase consumer trust in [company name] following a series of product recalls?
Graduate recruitment via the Internet	To what extent and in what type of context is Internet-based recruitment and selection of graduates effective and why?
Supermarket coupons as a promotional device	In what ways do the issue of coupons at supermarket checkouts affect buyer behaviour?
Challenger banks and small businesses	How has the emergence of challenger banks impacted upon small businesses' financing and why?

Your research question and research aim are complementary ways of saying what your research is about. However, neither gives sufficient detail about the steps you will need to take to answer your question and achieve your aim.

Table 2.3 Examples of research questions and related research aims	
Research question	**Research aim**
How effective is a media campaign designed to increase consumer trust in [company name] following a series of product recalls?	The aim of this research is to assess the effectiveness of a media campaign by [company name] designed to increase consumer trust following a series of recalls of its products
In which situations and to what extent is Internet-based recruitment and selection of graduates effective and why?	The aim of this research is to understand situations within which Internet-based recruitment and selection of graduates is effective and why
In what ways do the issue of coupons at supermarket checkouts affect buyer behaviour?	The aim of this research is to explore how the issue of coupons at supermarket checkouts affects buyer behaviour
How has the emergence of challenger banks impacted upon small businesses' financing and why?	The aim of this research is to explore how the emergence of challenger banks has impacted upon small businesses' financing and why

To do this you will need to devise a set of investigative questions or research objectives. Your research question may be used to generate more detailed investigative questions, or you may use it as a base from which to write a set of research objectives. Objectives are more generally acceptable to the research community as evidence of the researcher's clear sense of purpose and direction. Once you have devised your research question and research aim, we believe that research objectives are likely to lead to greater specificity than using investigative questions. It may be that either is satisfactory. Do check whether your examining body has a preference.

Research objectives allow you to **operationalise** your question - that is, to state the steps you intend to take to answer it. A similar way of thinking about the difference between questions, aims and objectives is related to `what' and `how'. Research questions and aims express 'what' your research is about. Research objectives express `how' you intend to structure the research process to answer your question and achieve your aim. In this way, research objectives can be seen to complement a research question and aim, through providing the means to operationalise them. They provide a key step to transform your research question and aim into your research objectives.

(Saunders, Lewis & Thornhill, 2016, pp.46)

Operationalizing research questions

For an empirical research question (the sort you can only answer through data) to be operationalizable (see also Cohen et al., 2007: 81-83), there must be a way of addressing it, in terms of identifying the appropriate data, collecting and analysing it (see the section 'Implications of your research questions for data, data collection and analysis'). Often there are indications of how to do this in the research question itself. For example, as we have seen, a research question like 'What reasons do French teachers working in UK primary schools give for the teaching of French to Year 6 primary school children?' suggests that the researcher would elicit data, for example, might ask teachers a set of interview (or questionnaire) questions which together, properly analysed, would address this research question. Note though that the words 'What reasons do [they] give ... ?' constitute an important reminder that we cannot get at people's actual reasons directly from what they say - at best, these are 'reported beliefs; the beliefs they 'give' (see also Edley and

Litosseliti, this volume). This has implications for the interpretation of findings and the strength of claims that can be made. (Note also that interview questions are not the same as research questions. It would be unreasonable to put your research question directly to a respondent.) Secondly, every 'content' word in a research question matters. To operationalize your research question, you will need to know exactly what each word is to mean as far as your research project is concerned (a 'working definition, that is, 'for the purpose of this study/dissertation/ thesis').

For the set of research questions given earlier about the `quotative use of be like', we would need to be clear about what we mean by this, and how to recognize such use when it occurs in the data. It may seem obvious (as in the example in the research question itself, He was like 'I can't stay here'), but there will be cases where be like is not quotative (e.g. 'I'm like my friend') and other cases where it is unclear, and parameters will need to be drawn. Also in need of a working definition is the concept, in these research questions, of adolescence (who counts as an adolescent?). The terms in your questions will also correspond to your theoretical and epistemological focus: this is very evident in words like ideology or discourse, but even the word beliefs in a research question indicates that you consider your research participants' understandings as important, interesting and epistemologically valid (see also Mason, 2002).

Mason (2002: 19) also reminds us that we should ask of our research questions, 'Would anyone but me understand them?' It is crucial that the answer is 'Yes' - especially if aspects of your study are to be replicable. If others cannot understand your research questions, it is worth considering whether they are, in fact, formulated in a way which is clear enough for you to address them properly.

(Sunderland, 2018, pp. 24-25)

Quantitative research is based on deductive reasoning. This means the design starts with a theory or hypothesis and is followed by evidence to support or refute this. Quantitative designs need a large number of participants to be able to perform the analysis, which is usually statistical. Statistical tests require a minimum sample size, that is, number of participants, for the results to be accurate. These tests are usually calculated using specific software, for example, Statistical Package for Social Sciences (SPSS). Analysis of data is often inferential, which means it is possible to generalise findings to the population beyond the sample tested. To conduct this analysis, it is necessary to have variables and a systematic operationalisation of these. Operationalisation refers to the specific questions used to measure a specific variable. This needs to be done before the data is collected and is referred to as an a priori design.

(Woodrow, 2022, pp. 70-71)

References

Andrews, R. (2003). *Research questions.* London: Continuum.

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design.* Cambridge: Cambridge University Press.

Burns, R. B. & Burns, R. A. (2008). *Business research methods and statistics using SPSS.* London: Sage,

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.

Cook, V. (1986). The basis for an experimental approach to second language learning. In V. Cook (ed.), *Experimental approaches to second language learning* (pp. 3-21). Oxford: Pergamon Press.

de Vaus, D. A. (2001). *Research design in social research*. London: Sage.

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford: Oxford University Press.

Gass, S. M. (2010). Experimental research. In B. Paltridge & A. Phakiti (Eds.). *Continuum companion to research methods in applied linguistics* (pp. 7-21). London: Continuum.

Hatch, E. & Farhady, H. (1982). *Research design and statistics for applied linguistics*. Cambridge, MA: Newbury House.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* Boston, MA: Heinle & Heinle.

Johnson, B., & Christensen, L. B. (Eds.). (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed., pp. 30–56). Los Angeles, CA: Sage.

Litosseliti, L. (2018). *Research methods in linguistics* (2nd ed.). London: Bloomsbury Academic.

Mackey, A. & Gass, S. M. (2022). *Second language research* (3rd ed.). London: Routledge.

Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.

Paltridge, B. & Phakiti, A. (Eds.). (2010). *Continuum companion to research methods in applied linguistics.* London: Continuum.

Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction* (2nd ed.). London: Bloomsbury Academic.

Saunders, M., Lewis, P. & Thornhill, A. (2016). *Research methods for business students* (7th ed.). London: FT/Prentice Hall.

Sunderland, J. (2018). Research questions in linguistics. In L. Litosseliti (Ed.), *Research methods in linguistics* (2nd ed., pp. 14-24). London: Bloomsbury Academic.

Woodrow, L. (2022). *Introducing researching English for specific purposes*. London: Routledge.