

361-208

Second Language Testing

May 7th - May 11th

Aren Tyr

61503641027

Organisation/Table of Contents

- ▶ The various tasks have been colour coded and organised by a tab down the left-hand part of the slide as follows. Each section is also individually numbered.
- ▶ Click on the section links below to jump straight to the corresponding slide when in slide-show mode:

Task	Task/Section	Slide Number	Colour
1.1	<u>Testing Basics</u>	4	Purple
1.2	<u>How do they know</u>	13	Yellow
1.3	<u>How to write test items</u>	23	Blue
2	<u>Error & Correction</u>	33	Green
3	<u>Writing Performance Outcomes</u>	62	Yellow
4	<u>Assessment Based Language Education</u>	74	Red
5	<u>Testing Skills Input (Jigsaw Group Work):</u>	86	Brown
5.1	<u>Testing Listening (Group 1)</u>	87	Brown
5.1	<u>Testing Listening (Group 1) Critique/Notes</u>	101	Brown

Task	Task/Section	Slide Number	Colour
5.2	<u>Testing Speaking (Group 2)</u>	102	
5.2	<u>Testing Speaking (Group 2) Critique/Notes</u>	146	
5.3	<u>Testing Reading (Group 3/My group)</u>	147	
5.3	<u>Testing Reading (Group 3/My Group) Critique/Notes</u>	155	
5.4	<u>Testing Writing (Group 4)</u>	157	
5.4	<u>Testing Writing (Group 4) Critique/Notes</u>	186	
5.5	<u>Testing Grammar (My work)</u>	187	
6	<u>How to validate tests</u>	211	
-	<u>Action Reflection/Self Reflection</u>	232	

Task 1.1 – Testing Basics

- ▶ 1. Why test? 5 main purposes for testing
- ▶ 2. Different test types?
- ▶ 3. Validity, reliability, and authenticity

1. Why test? 5 main purposes for testing

There are actually many different motivations behind testing, many of which have their roots in complex socio-political and economic factors. The attempt to quantify something – whatever it is – is an attempt to make it ‘known’ and classifiable. If you can quantify something, you can put it against a scale with the implication of a hierarchy. In short, then, testing and power are deeply inter-related.

Such themes, all of which invite multiple book length explorations, fascinating as they are, are outside the scope our interests for this module, so instead we shall selectively focus on a ‘narrow’ definition of testing: testing is an instrument for assessing, in an objective manner as is feasible, what has actually been taught and successfully received by our students. How much of our teaching ‘input’ have they been able to convert into ‘output’.

Essentially then, testing is a practical tool for evaluating how well our students are performing on the basis of what we have managed to teach them.

If we now look specifically at second language acquisition, we can state five broad purposes/objectives for our tests:

Achievement – Discover whether students know specific language taught over a series of lessons, i.e. How much learning have they done of the taught materials.

Proficiency – Determine the student's overall general level of ability in the target language (typically against a worldwide index).

Achievement & Proficiency – Identify what to actually teach on a course going forward, by combining the two goals listed above: discover what the general ability of a representative student from your class/course is, and also determine how much of what has been taught thus far they have successfully learned.

Diagnostic – Establish whether your students have knowledge of a particular language feature, or features, under investigation.

Placement – Determine which level or grade to place new students on in an existing series of courses. Placement tests are extremely valuable for private language schools; e.g. if someone has already studied German seriously and methodically for two years, it makes no sense to start them at an absolute beginners level, as they will simply be wasting time repeating material they already know well.

2. Different test types; criterion referenced & norm-based.

The different types of test strongly correspond with the five main purposes of testing detailed above.

Achievement tests – These are ‘non-global’ in the sense that they focus almost exclusively on material that has been previously taught in the particular course/context in which they occur. In other words, the testing content is constrained and pre-determined by the syllabus. The objective is to determine what the student has learned on the basis of what they were taught.

Achievement tests should be, or belong to the category of, ‘**criterion referenced**’ tests. Here, *criterion* refers to a metric against which we assess what the students can ‘do’ and under what circumstances they can do it. Such criteria are pre-determined *a priori* before the course and essentially form the foundation for our planned performance outcomes for students on the course.

Proficiency tests – These are ‘global’, in the sense that by definition they will contain unknown language, and are used to determine what the student actually knows (in general across the target language).

Unlike achievement tests, proficiency tests are instead ‘**norm based**’. The main feature of these is that they are used to specifically ‘rank’ students on a scale. In language education, this is used to group students with a similar overall level of ability into the same class. Such norm based testing is often heavily involved

with machinations of power and politics alluded to in slide 4 above, in a vastly complex number of ways.

Diagnostic tests – Focussed on determining whether the student knows about specific language features and their meaning. So here, they will typically be criterion referenced where the set of criteria specifically correspond to the language features we wish to ‘diagnose’.

Placement tests – Used for determining correct placement of a student within a pre-existing series of courses/modules/levels based on their current knowledge/proficiency.

This would naturally lead one to assume that placement tests are simply proficiency tests, albeit perhaps more course/context specific than proficiency tests. However, in general, language proficiency examinations are more ‘closely proctored, more comprehensive, and rigorous’, usually reflective of the fact that they are typically used for determining admittance onto fairly exhaustive programs (e.g. a language degree course in the particular target language, or, very commonly, for ESL learners from non-English language countries who wish to study for a degree which will be taught in English). Placement tests, therefore, can usually afford to be more lenient and less rigorous, since they are orientated towards classifying the student into a level at which they will be able to manage and progress with their learning, rather than intensively scrutinising their ability in as much detail as is feasible from a highly critical ‘academic’ language assessment perspective.

Aptitude tests – These are actually more of a ‘meta-test’ of sorts. Rather than test for particular knowledge or ability regarding the target language, instead these are focussed on evaluating the capacity of the student to actually study and learn languages in general. Effectively, then, this is more about testing their current cognitive skill-set, and specifically so in regard to their potential language acquisition capabilities.

3. Validity, reliability, and authenticity

Validity – What is the correspondence between what the test actually measures and what it is *intended* to test? Whilst it would be ideal if these two were always identical, in practice sometimes a test may actually end up evaluating skills/capacities other than what was intended. Frequently, indeed almost always, a test will end up testing skills/capacities in addition to what was primarily intended. This is due to the fact that language skills are closely intertwined and interdependent. To give one example, there is usually a fairly strong relationship between reading ability and writing ability.

With validity, we can consider it from two angles: *content* validity, namely, does the test actually evaluate the skills that it was intended to test (i.e. Does the test taker have to actually produce them during the test), and *construct* validity, does the test accurately evaluate all the features that are used to constitute that particular construct. A well designed oral examination would

test grammar, pronunciation, vocabulary *and* fluency, since all four of these are characteristic qualities that we would associate with any competent native language speaker in the target language. If the oral test does not then evaluate all four of these qualities/aspects, we would say that the test has poor ‘construct validity’ since it does not accurately test the target construct in this case (i.e. fluent speaking in the target language).

Reliability– Can the test be supported or corroborated with other data? Do we have other assessment devices/measures against which we can check to see how consistent the test performs across various cohorts?

If a test continually generates results that present a strong discrepancy with the students observable classroom performance and the teacher’s intuitive sense of the various students ability (an experienced teacher normally develops a pretty good sense as to how well a student is progressing, particularly if they have the opportunity to teach and work with them over an extended period of time, e.g. a semester or academic year), that would suggest there is something highly suspect with test, that it is likely to be an unreliable measurement device.

Similarly, more broadly, we would expect a well designed *reliable* test instrument to yield broadly comparable result ranges across entirely different classes/institutions, assuming the material/syllabus was taught to at least a largely similar level of quality and quantity. A test that cannot serve this purpose would be no use if we wished to deploy it for proficiency testing, for example. This property is its *replicability*.

Authenticity – What is the mapping between our test and the real world? To what degree to it simulate or actually bear a relationship to the skills/performances required for real-world usage? In the case of second language acquisition, for example, if someone tests at CEFR C1 or C2 level we would expect them to be able to go the country of the target language, and be able to confidently use the target language to achieve day to day tasks and successfully use it to work professionally (certainly at C2!). If that was not the case, then we would have strong reason to suspect the authenticity of the C1/C2 metric as a result of their language tests.

Of course, we need to keep in mind that the almost unavoidable ‘artificiality’ of the test environment means that the test is always going to be inauthentic to some degree, but it should at least test similar skill sets. For example, if a student could select correct items on a series of multiple choice exam question which closely assessed meaningful/representative dialogue of ordering train tickets from the ticket counter, it is likely they have at least got the correct conceptual understanding of the language features in question. An audio dialogue, meanwhile, which could also be used and might *seem* to offer a perfect test substitute, would nevertheless still suffer from some degree of authenticity since in the real world there are endless background noises, the endless individualistic quirks of isolated native language speakers, the speed of speech and their prosody, the use of colloquial language, and so on and so forth, none of which is likely to be captured in a typical perfectly recorded and classically ‘grammatically perfect’ audio clips, spoken with moderate or slow speed used in prototypical language testing environments.

So instead we have a ‘compromise’ or ‘level of acceptability’ when it comes to authenticity. Testing a language feature using multiple dimensions can help to at least partially overcome some of these weaknesses; e.g. You could test using an audio clip/recorded dialogue in the first part, then in the second use a multiple choice question set to evaluate the student’s knowledge of the same underlying language feature/context, to give you a more meaningful insight into the *likelihood* of them being able to apply their knowledge/skill of that language feature in its real world authentic counterpart.

We have to recognise that short of conducting a ‘real-world’ protracted assessment, of a type that is infeasible and too slow/expensive, all testing/assessment does suffer from being ‘completely artificial’, to paraphrase Noam Chomsky. If we could test second language assessment by following a student around for two weeks ‘big brother’ style in the real world language environment where they are forced to use the whole host of language skills under assessment across a series of actual real world demands of that language in its context and cotext, it is fair to say we would likely be able to get an extremely powerful assessment into their language proficiency and capability with the target language. Such a scenario is mostly fanciful, however, as it would simply not be affordable to test in such a demanding way, so instead we make do with less authentic and less accurate testing methods.

Task 1.2 – How do they know?

1. How the students knows, and how this affects testing.
2. A few notes regarding memory
3. ‘How’ do they know?

1. How the students knows, and how this affects testing.

a) Learning modality – methodology of instruction & assessment

The four skills of reading, writing, speaking and listening all require specific work in order to develop. Though there is a strong relationship of development, it is nevertheless perfectly possible (and indeed happens) to develop them asymmetrically. Chinese ESL students, for example, who are known for their work ethic from their education system, tend to often have good reading and writing skills from their extensive textbook work, but are far weaker in speaking and listening in real world use due to the lack of native English language tutors able to provide suitable context and modelling, and above all and opportunity for for spoken language and conversation. Passive viewing of onscreen video is insufficient to compensate for the lack of productive use of spoken language to really develop both speaking and listening, since a conversation is a dynamic exchange that requires both skills. Textbooks, and written English text, by contrast, is almost universally available and essentially often free via the internet. This feature is compounded by their personal emphasis on written English for the purposes of entrance to academic institutions in higher education who deliver their courses and materials in English.

Testing therefore needs to take account of learning biases such as these that may exist in ESL learners. The test needs to appropriately reflect both the content and the method of delivery of the taught material, particularly if we are talking about *achievement* tests. The example I have cited regarding some Chinese ESL students above, is a representative case of the difference between *proficiency* and *competency*; no doubt they will still have had to pass

Some oral examinations as part of their testing regime; however, such testing does not in itself necessarily translate well into real world use due to the myriad additional difficulties that English language use in context presents (e.g. Speed of speech, variety and range of accents, dialects, style of speech, use of idiomatic, colloquial and slang phrases, ambient background noise, patterns of intonation and stress, etc.). In this case, there is no substitute for exposure to as much 'real' world English as is possible.

b) Memory vs. development – structure vs. agency

Classrooms and textbooks are excellent for the formal presentation of grammar, and for the explicit presentation of language features. However, herein lies a potential difficulty, in that native speakers actually acquire most of their language potential inductively through exposure to the language in real life contexts. In short, they develop a significant proportion of their language through *functional use*. A child soon learns that in order to have its demands and requests met, it must explicate these demands through language. Physical body language may go some of the way towards achieving this objective, but once their cognitive abilities have developed further, and their corresponding desires become both more complex and abstract, the communicative demand for language becomes overwhelming compelling. This is further reinforced by a 'reward' mechanism; if I don't ask for something, it is highly likely I won't get it, unless by blind luck or habit

the adult happens to get the item in question. If I am able to express my 'want' in language, I have a chance of getting it. With such a motivational 'carrot' dangling throughout our emotional-cognitive development years, we have here a case of 'agency' driven language acquisition. I have a purpose for my language, so I will make use of it.

Textbook language learning by second language students has a radically different dynamic. Instead of by directly being driven by agency, a lot of the time the progression of language is from a top-down structural approach, analytical in character. One of the problems with this is that it is likely to lack the deep embedded contextual 'richness' of language acquired naturally by 'use' and by the demand to express needs/wants that you would like met, together with the deep set of emotional links that 'root' all these language exchanges. These 'embedded' emotional-contextual links help to ground memories so that language features become permanently retained, since they are 'real' in a way that textbook examples tend to lack.

C) Learning context – numbers of students & frequency. Classroom vs real life.

The context relates to issues of memory and development described above. Issues related to the artificiality and limitations of classroom teaching have also been described above. So we then have the rather more basic issue of language exposure/time spent learning.

Even if we accept that a child may be not exposed to language for many hours of each day (sleeping, playing silently, running around in the countryside, staring at the ceiling, eating, etc.), nevertheless, by the age of 5 it is likely they could have had 5000 hours of exposure to English language or more. Even with their neural-cognitive limitations as a baby, this goes a long way to explain the potential wealth of vocabulary even a young child may possess, versus a fully developed adult language learner who after studious effort over months or years might still only know a few hundred words, far less than the child. The difference merely continues to accelerate until by the time the teenage years are reached, the richness of language the native language speaker possesses seems almost impossible for a second language learner to obtain.

Yet how much of this is mere quantity? A *very* dedicated language learner living outside the target language country may perhaps manage to devote 1-2 hours per weekday to a language, and normally always in an artificial context. That's something of the order of 500 hours *per year*. A native speaker, by contrast, will be getting typically at least 8 hours or more of language exposure per day, every day, if we assume they go to school and interact with parents at least somewhat outside of school hours. At weekends this will also be the case. So that is probably somewhere in the region of 3000 hours per year, again, conservatively. Nearly all of this is real world language use in context. So the second language learner will necessarily have a far slower rate of development. If we then factorise this by a more realistic scenario, whereby the language learner has two classes of

one hour per week, plus does one hour of homework, for a total of 3 hours per week during term times, then the figure could have dropped to something closer to 250 hours per year.

Estimates by the FSI in the US variously suggest between 1500-3000 hours are typically required to reach a reasonable degree of proficiency in a language. The exact total – notwithstanding endless issues of quality, context, and relevance – is actually immaterial here. The point is, regardless of the number, it is incontestably a very large number of hours, as evinced by the fact we have a planet full of people who find learning another language an extremely difficult and time consuming affair. So even with better learning strategies and teaching, a miracle is not on the horizon (short of a implanted brain chip!). This being so, a language learner who is devoting 250 hours per year to a language will probably take a decade to get even to a modest standard. Some university students who do a language degree can reach something approaching “C2” standard by the end of their degree, but this likely the result of a significant number of hours study. They perhaps manage to get something closer to thousands of hours over the course of their A-Levels, university degree, holidays, trips and general persistent exposure to media from their target language due to a devoted interest.

d) Social context – attitudes towards ELL

The ‘devoted student’ described above connects with the topic of motivation. Given the large degree of persistent effort required for SLA, the learner’s motivation is key if they are to sustain that effort for the extended period of time required for them to reach proficiency. Do they actually make use of the language to develop it? Fear of imperfect utterances can paralyse a language learner as soon as they encounter the language ‘in the wild’ of the real world. Without using it in a real world situation, they will forever lack the embedded context described above, to complete the cognitive ‘circle’ of learning.

Do they have any sort of support network? If learning that language is of no value to their society, or indeed anyone except the student, effectively they become a lone ranger, solely dependent upon themselves to sustain their efforts. They have no one to support them, and will likely lack all the socialisation and scaffolding that can be so essential for developing their language further. They have their classroom, if they attend one (versus learning online, solo, by books, etc.); but again, how many hours per week is that? Generally very few.

2. A few notes regarding memory

The structure of the human memory system is still extremely poor understood, and hotly contested, Nevertheless, despite all its limitations, the conventionally accepted model of 'short', 'working' and 'long term' memory, rather like a set of metaphorical conveyer belts, is methodologically useful.

Short term memory (such as in a language learning lesson, where you are exposed to new language items) is relatively fragile and limited. The accepted wisdom is something of the order of 6-10 new words per lesson can be taught. By the same token, instruction of new grammatical structures and complex syntactic information should also be introduced slowly, in a piecemeal nature. This information generally needs to be recorded immediately in some manner so that it can be deliberately reviewed, otherwise it will likely be completely forgotten within days.

These new items need repetition and repeated exposure, bringing them into your 'working' memory which is rather like the RAM of your computer. This information needs to be assembled into some sort of meaningful structure or pattern for the learner, otherwise they will simply remain empty lexical items that ultimately get pruned away as meaningless, This working memory is characterised by *conscious effort*.

Repeated *use* of these structures in working memory eventually commits them to long term memory where they are almost permanently retained. Even after many years, they will likely not be completely forgotten, though they might perhaps need ‘refreshing’ to bring them back up to the level of automatic and effortless recall.

Long term memory therefore tends to usually have this unconscious property, such that that items and structures within it can be used with no explicit conscious thought. This is where we really want our language acquisition to end up, because this is the only place where the production/reception of language can happen quickly enough to enable fluency/comprehension when speaking or listening at native language speeds (or to enable the production of written text, for example, without taking a painful amount of deliberate effort to type even the most rudimentary of sentences!).

3. ‘How’ do they know?

I discussed above the example of the discrepancy often apparent between Chinese ESL students reading/written language and their spoken/listening skills.

This was a case of their language learning being specifically tailored for the

purposes to which they most need it – prioritisation. If we consider three cases:

- i) Cognitive knowledge of language through written testing
- ii) Ability to listen, read, and respond to test questions
- iii) Ability to write an academic essay

Point i) would often relate to their ability to re-present explicitly taught language structures (e.g. formal grammatical syntax, for example), ii) would be a frequent demanded language skill for academic work and a mainstay process in language education, whilst iii) meanwhile is our gold criterion, as almost every subject in higher education requires such long-format language presentation. Though i) might be comparatively easy and seductive to test, and ii) will probably happen as a matter of course during the process of teaching the language, it is really iii) that is likely to be a strong focus or objective for our students.

Writing an essay places a very strong demand on written language production, as it an exceptionally ‘high level’ language skill, due to the requirement to construct coherent and concise description, analysis and argumentation, so we can see here why the bias towards reading/writing in many ESL classroom early on tends to occur as a pragmatic principle of necessity if students are ever to reach the required standard by typical university ‘age’ (notwithstanding the increasing trend of ‘older’ learners now attending university too).

Task 1.3 – How to write test items

A multiple choice test for grammar¹ with 5 test items

1. A few notes on my test
2. The Test

¹*I have interpreted 'grammar' here to mean grammar in its wider more general use of 'language' in general; in any case, the test I have designed is focused on detecting comprehension of nuance related to vocabulary and meaning as a result of some comparatively complex grammatical forms.*

1. A few notes on my test

After some consideration as to the manifest difficulties involved in constructing a multiple choice test, I decided I would use a 'scenario' that links all the questions so that they occur in a narrative. The test is for fairly moderately advanced level language learners, as it requires them to intensively read the scenario and the information presented, then carefully consider their answer based on typical collocations of language form and use, together with the overall meaning of the phrase,

I did a trial run of this test on my Thai girlfriend, who is reasonably advanced at English (I would guesstimate around B2), and she was caught out by question 4 and somewhat by 2. I considered revising question 2, as a problem with writing distractors is a decision has to be made as to how explicitly 'wrong' you want to make them; make them too wrong and they are no longer a distractor, make them only fractionally wrong and you invite a philosophical exegesis as to why the proposed multiple choice question answer is not necessarily exclusively correct. A problem that can occur with multiple choice questions, then, is that it can be often be the case that 'it depends'. It 'depends' because no use of language is ever entirely unambiguous.

This 'it depends' characteristic also highlights the difficulty of writing good test items. You want them to be as unambiguous and clear as possible. Yet you want your distractors to be tempting and 'plausible'.

Overwhelming evidence to this point exists in the fact that language must always be *interpreted*; it is not simply 'passively' received. Naturally, tests at a low level tend to avoid this issue as they can make use of very brash distinctions that invite little to no ambiguity, or simply test grammar in a pure unfettered form, so the distractors in such cases can be stated to be wrong in a very 'pure' sense.

In any case, after some consideration, I am comparatively happy with my attempt, though I recognise that as we have discovered during this module, in order to ensure the test items are truly *valid* they must be subjected to a long process of review and amendment.

My test is a **general proficiency** test. I am interested to detect whether they can pick out some reasonably high-level collocations, some grammar specific features, inferences based on correct language comprehension/meaning, and whether they can intensively read to extract critical information to determine the correct answer. The test may suffer from trying to test for too many things all at once, even as a proficiency test; it is difficult trade-off when trying to test moderately advanced language learners, and I have never written a multiple choice test before.

2. The Test

Example Question (and model answer)

You step into the street. There are grey clouds overhead and now it has just started raining very heavily. Your friend steps out your front door with an umbrella and asks you whether you want to take it with you.

You decide you want it, so what would you say?

Select the most appropriate answer:

- a) “Yes please, I can take those.”
- b) “Yes please, it looks like it is going to rain.”
- c) “Yes please, I had scrambled eggs for breakfast.”
- d) “When can I take your umbrella?”

The correct answer is **b**).

- a) Is incorrect as ‘those’ is the incorrect pronoun for the single countable noun umbrella.
- c) is incorrect as there is no logical relationship between scrambled eggs and the umbrella in this context.
- d) is incorrect as your friend has already just offered you the umbrella, so ‘when’ is not applicable.
- b) Is correct as you are accepting the umbrella, and have commented on the fact that it is raining, which is logically related since it is the usual purpose of having an umbrella.

Question 1

You get into the taxi. The driver says:

“Where would you like to go?”

Select the most appropriate response:

- a) “I would like to go there.”
- b) “The trains are at the station.”
- c) “To the train station, please.”
- d) “Would you like to take the train?”

Question 2

You arrive at the train station. You need to buy a ticket but do not know where the ticket office is.

You notice a woman who has a pair of train tickets in her hand whilst putting her purse back into her handbag.

Select the most appropriate question.

- a) “Excuse me, where did you buy your train tickets?”
- b) “Excuse me, can you tell me about train tickets please?”
- c) “Excuse me, what is the destination of your train tickets?”
- d) “Excuse me, where is an office please?”

Question 3

You reach the ticket office. You try to buy a ticket to Birmingham. The ticket officer tells you the ticket will cost £55. You then realise you do not have enough money since you only have £40 in your bank account.

Select the most appropriate response:

- a) “I have £55 in my account; I will give you £40, thank you.”
- b) “Since the ticket is £40, you can take the money from my account. Thank you.”
- c) “I only have £40 left in my account. Could you sell it to me for £40?”
- d) “Thank you, you will take £40.”

Question 4

The ticket officer has sold you a ticket to Birmingham for £40 on a special offer. You reach your platform, number 8. Your train is due in 10 minutes time. A train manager from the train company comes onto the platform and says:

“The train arriving here on platform 8 heading to Birmingham has been delayed by 25 minutes. Network Central West apologises for the delay.”

What does this mean? Select the correct answer:

- a) The train will be arriving in 25 minutes time.
- b) The train will arrive 25 minutes later than it was scheduled.
- c) The train will arrive in Birmingham in 25 minutes time.
- d) The train arrived from Birmingham 25 minutes ago.

Question 5

The train arrives on the platform and you get on. Your seat reservation says you are in seat B2. You walk down the carriage to seat B2, where you are supposed to be sitting. A man is sitting there talking on his mobile phone. You tap him on the shoulder and speak to him.

Select the most appropriate statement to say to him:

- a) “Please sir, would you kindly stay with seat reservation B2 here.”
- b) “Please sir, but my seat reservation is yours. Is mine B2?”
- c) “Excuse me sir, but your reservation is seat B2. I have a seat.”
- d) “Excuse me sir, but this is my seat. I have a seat reservation for B2.”

Answer Grid

Question	Correct Answer
1	c
2	a
3	c
4	b
5	d

Scoring Weight

Question	Percentage
1	15%
2	20%
3	20%
4	25%
5	20%

Task 2 – Error and Correction

1. Error as ‘learning’
2. An error or mistake?
3. Looking for what’s right
4. Examples of inter-language
5. The language learning curve & inter-language
6. Learning stages/Categories
7. Socialisation & scaffolding
8. Mediating factors between ‘input’ to ‘output’
9. Order of Acquisition
10. Correcting inter-language mistakes
11. Tactics for correcting inter-language mistakes
12. A commentary on correcting in general
13. The learning paradigm

1. Error as 'learning'

If a language student speaks or writes in a grammatically incorrect manner, is that due to ignorance (i.e. they are either not *aware* that it is incorrect, **or** they entirely *lack* the requisite knowledge to produce it in any other manner), or is it due to an unintended slip (i.e. If their attention is drawn to it, they can self-correct and modify their output)?

This distinction is what separates an **error** (they have no possibility to self-correct) from a **mistake** (they can potentially self-correct). Separating these two types of incorrect language use is important since both require different remedial strategies from the teacher.

In an idealised scenario, the resultant 'inter-language' that is a characteristic of learners who produce errors and mistakes would not occur; we would instead simply progress from 1.) Comprehensible input, to 2.) Reinforcement & Practice, to 3.) Functionality (i.e. Output) in a linear manner of increasing complexity. Rather like progressing from start to finish of a language textbook, we would start on page 1 and then simply work through page after page, unproblematically, until we reach the end of the book (or series of books) having mastered the language. Unfortunately this almost never occurs (if someone could progress in such a manner, we would probably consider them a savant, or a linguistic genius).

So what we have instead, then, is students who produce a vast quantity of partially correct utterances; some elements are completely wrong, but some elements are *right*. As a result, what they produce is still meaningful, even if it is not yet completely correct. It is this production of inter-language that is actually indicative that learning is taking place.

2. An error or mistake?

Whether the student has made an error or a mistake is essentially relative to what they have thus far been taught. An error is an unavoidable feature of language learning, where the learner is attempting to construct a piece of language that exceeds their current knowledge of the grammar; a mistake is a lapse in concentration or misapplication of previously taught language material. In theory it is avoidable, but in practice no student would ever meet such a criterion of perfect studiousness, since we cannot necessarily expect a language learner to be able to dedicate hours every single evening practicing when they undoubtedly have other pressures on their time. Not to mention other externalities such as sleeping poorly, emotional distress, and other factors that interfere with the learning process, etc. Given these constraints, then, some mistakes will naturally occur. Typically a teacher should generally correct most *mistakes*, depending on the nature and goals of the activity (e.g. If the activity is an open speaking activity where *fluency* is our primary target, then continually interrupting the student to correct mistakes may be highly counterproductive).

Let us consider the following sentence:

“Please wait here before getting at the bus”

Here we have an incorrect use of the proposition *at* instead of *in*. If the learner in question has received lessons on the use of these propositions (i.e. *on*, *at*), it is a mistake. If we have yet to fully teach all the propositions, then it is an error, and shows

That the learner is hypothesising about new language constructions.

A common error in English learning relates to the word order used in questions, which can take a fairly complex grammatical form. For example,

“Do you know what are you going to eat?”

Would be a fairly prototypical example, since the learner does not realise that the word order changes compared to a more simple question. Here the question is indicative of being slightly above the level of the learner, so further tuition is required; it indicates that learning is taking place, however, as the student is attempting to formulate new constructions based on extrapolations from their current grammatical knowledge (and perhaps by drawing inferences from their L1 grammar knowledge).

3. Looking for what's right

Whenever a language error/mistake occurs, it is natural to immediately focus on this language defect. In doing so, we tend to overemphasise the errors/mistakes students make and not give enough credit to the many things they have correctly executed in order to construct their language utterance. A more balanced approach would be to highlight these positive aspects as well as the error/mistake.

Some examples of “what's right” in the following example phrases/utterances:

1. *Mummy... biscuit!*

This sort of utterance from a baby/toddler actually demonstrates a logical understanding of the fundamental communicative intention behind language and relationship between sounds and objects. The toddler has identified its mother with the noun 'Mummy', the edible food item with 'biscuit' and has issued an imperative that is unambiguous in meaning: "Mother, give me that biscuit now!".

2. *He came here since yesterday*

The word order is correct, as is the SVO structure (subject = He, verb = came/[was]/[existence/presence], object = yesterday). The learner has also identified that 'since' is used to refer to a particular moment in time. They have simply not realised that when combined with yesterday it takes on a continuous form; so that it is the incorrect verb form of the present perfect continuous tense, where it should have been 'has been' instead of 'came'. What they probably meant was instead a past simple utterance, in which case they have simply added the additional unnecessary 'since' which just needs deleting to form a correct sentence. So it is an example of incomplete grammatical knowledge, but largely correct in application.

3. *What do you like to do today?*

The vast majority of this sentence is correct and shows good grasp of question formation in English grammar. It is correctly WH fronted, 'do' is correctly ordered/reversed in the two different applications in the sentence (with 'you' and 'to'; this is quite a difficult feature of English), and the vocabulary and subject/verb agreements are all correct. It is simply missing the one word 'would'.

4. *Do you still in Saudi?*

Again, most of this sentence is actually correct. The sentence is correctly fronted, with 'do' and 'you' ordered correctly. They have selected the correct adverb 'still', and also selected the correct preposition 'in' to go with the location Saudi. Recognising that Saudi is a proper noun referring to this designated country, they have correctly capitalised the noun. The only word missing is 'live', the verb that goes with the adverb 'still'.

5. *Where are you exercise?*

Correctly WH fronted, and 'are' and 'you' correctly ordered. The verb exercise correctly matches the subject word 'where'. The only error/mistake is the incorrect form of the verb, missing the '-ing' form to indicate present continuous.

Alternatively, if they meant it to ask where we exercised having presumably completed such exercise, as a past simple sentence, then they used the wrong auxiliary 'are' instead of 'did'. In either case, it only one small error in an otherwise correct sentence. We would require more context to this sentence to identify in what sense it was intended to specify which of the two errors was made; both are nevertheless quite plausible and prototypical errors.

6. *What you favourite dessert?*

Correctly WH fronted once more, and in correct SVO structure. A classical case of not applying the complex verb 'to be' by placing 'is' and changing you to the possessive personal pronoun 'your'. The general meaning of the sentence is correct, as any English native speaker would be easily able to infer the missing 'to be' verb and modification of you to 'your'.

7. *Teacher what mean?*

Correct SV structure, fronted with the subject (teacher) and verb (mean). The student has mostly correctly formed the interrogative; it is simply missing the auxiliary 'do' and pronoun 'you'. The verb mean is correctly paired with a suitable subject.

8. *Me present teacher*

Assuming that they are here describing having given a gift to the teacher, then the utterance is missing the verb 'give', preposition 'to' and articles 'a' and 'the', together with incorrect use of personal pronoun. Despite all these errors, they have nevertheless correctly followed the SVO structure, and identified the correct noun for the object 'present'. The sentence is still a meaningful piece of inter-language.

9. *Tell me what are you doing?*

This sentence is only very marginally 'incorrect'. Its only real error is redundancy; the 'what' question as an interrogative already implies the instruction to the question receiver 'to tell', so explicitly restating the implied property intrinsic to the grammatical structure is merely redundant. Nevertheless, sometimes someone may wish to deliberately make use of this redundancy to add emphasis to the question receiver that they must give an answer, such as a parent to a child. The question then effectively becomes a command that must be responded to:

"Tell me, what are you doing?"

10. *The house was besides the river.*

This sentence is correctly ordered, with the correct use of articles. The correct verb ('was') has been used that matches the preposition. The SVO structure is correctly represented. The only error is the incorrect use of the proposition beside, where the addition of the one letter 's' changes the meaning entirely.

4. Examples of inter-language

Inter-language, therefore, is the production of partially correct English. It occupies a middle ground between meaningless random symbols and perfectly formed English sentences; generally, a lot of inter-language can be successfully understood by a native speaker as often there is a enough material that is coherent to successfully convey the basic message.

Some examples:

An cat was sitting on the porch.

Misuse of the articles 'a' and 'an' is common. This error is so minute that though it makes the sentence grammatically wrong, semantically it is essentially perfect. There is no ambiguity whatsoever as to what the sentence means. So this is only 'just' inter-language. Another error in this class of formally incorrect, but otherwise semantically correct sentences concerns the ordering of adjectives:

He cycled on his metal orange new bike.

When instead, it should be *He cycled on his new orange metal bike*, since we have an order of precedence for multiple adjectives. There is no semantic difference between the two sentences, however.

Examples concerning incorrect word order in questions has already been discussed above.

Another class of error concerns inter-language that is produced as a result of attempting to diligently following English grammar rules, which the expectation that the language is consistent, has a clear and coherent set of rules in the first place, and does not violate any of its own rules! Sadly, English does not meet any of these standards hence why English grammar is actually a set of conventions or accepted 'best-practice' norms rather than strictly defined rules.

Consider the word 'fun'. A common error I have spotted among Thai speakers when using English is with the adjective form of this word. It is quite reasonable to expect that if you describe something as 'funny', then what you mean is that 'it had the property of being enjoyable'. Unfortunately, the word 'fun' has two distinct (though often complementary) meanings, and the adjective 'funny' only refers to one of them.

[referring to a photo of Songkran water fighting]
Looks funny!

Though they may possibly be intending to mean that it generates mirth, in reality the prime meaning they are likely after is to describe it as looking like it is/was enjoyable. This is an example of a subtle inter-language error due to the variances in mappings of meaning between word-roots and their derivations. These mappings are largely arbitrary so must simply be memorised in use. Adjectives usually take a word root and add a suffix. Here, their intended adjective is in fact identical in spelling to the noun 'fun', viz. '*Looks fun!*'.

There are an almost infinite number of inter-language errors. But they all follow

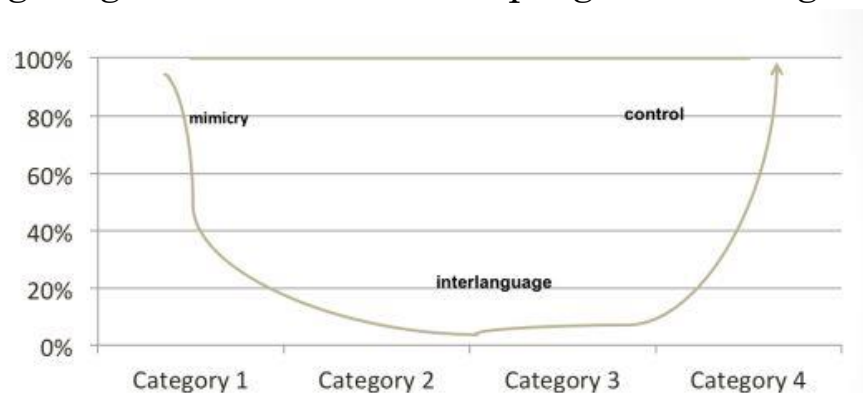
the same general overall structure, related to the acquisition of language. In very broad terms, a learner will first start out by either rote-memorising stock phrases (with no comprehension) that they can either speak or possibly write, before gradually progressing onto a increasing lexicon of individual 'tokens' that associate words with concepts or object (i.e. lots of nouns). They can simply give meaning to utterances by stating the nouns in some sort of relationship supported either by gestures, miming, or other basic nouns (e.g. 'water' + 'drink' {mime drinking} = [in a restaurant to a waiter] "Can I have some water, please?"). Next, whilst this process of vocabulary expansion is continuing, some rudimentary use of grammar structures will start to emerge, such as in functional language use in the form of questions. Various words will be missing and the word order may be incorrect, but some understanding of the most general language structure features (e.g. SVO order in English) may start to emerge. These then get progressively refined, until all the main basic grammatical structures start to be comprehended. The remaining part of the process is the long and gradual evolution of vocabulary, with an increasing comprehension of nuance, tone, style and increasingly rich application of the grammatical structures into more complex compound forms.

5. The language learning curve & inter-language

So, what we have is a 'learning curve', where learners progress from blind mimicry (rather like a parrot that can repeat certain words, but with little to no comprehension of their meaning), to the increasing production of 'inter-language',

which consists of various correct features of the target language mixed in with a host of incorrect features. Such inter-language may still be comprehensible to the native speaker depending on the severity and type of errors encountered; e.g. Minor or even moderately severe syntactic errors can often still yield meaningful output in terms of conveying the basic meaning or intention behind the utterance, yet even extremely minor errors in vocabulary could render the fundamental message completely unintelligible¹.

So we can state that most of the language learning process consists of a long period of the production of an inter-language. Lightbown & Spada represent this learning process diagrammatically with an accuracy curve of language use plotted against four 'markers' or leaning stages that all learners progress through in sequence:



¹This happened to me whilst attempting to buy washing up liquid in Big C; though the context of being in a shop, in the relevant household section, meant that it was fairly obvious as to the fact I trying to purchase a domestic cleaning product, so no real functional language production was even required, my complete ignorance of the Thai expression for 'washing up liquid' (for dishes and plates) meant that I had to go through a series of unsuccessful mimes, Google Translate attempts, and Google Image searches on my phone before I eventually communicated my intention accurately. So here we have an example of how *one* word can have a crippling effect on intelligibility, long before we consider the complexities of language in use as a *grammar*.

6. Learning stages/Categories

So what can we state in general terms about these categories? We can see an immediate relationship between the four categories and the schema of Gass & Selinker's SLA model. Recall that the integration process was iterative, and that with progressive iterations of exposure to comprehensible input that is successfully taken in ('intake') the language that has been 'integrated' steadily grows. This integration expansion broadly shifts stepwise through the four categories.

Category 1	Category 2	Category 3	Category 4
<ul style="list-style-type: none"> •Mimicry •No real comp •Q&A memorize •Non generative 	<ul style="list-style-type: none"> •Chunking •Wkg Mem + •Routines •Contextual trees •Vocab stockpile 	<ul style="list-style-type: none"> •Linking •Structure+ •Emergentism* •Context switching •detouring 	<ul style="list-style-type: none"> •Control •Generative •Automatic •Confident

Category 1 is our 'parrot' stage: With no real comprehension, we can simply replicate the sounds, or scrawl the exact same shapes (I could easily copy some Japanese Kanji in a pictorial manner without having the slightest comprehension at to what I had written), or memorise rote-formulas. A lot of travellers use this approach with some success to memorise the sounds of straightforward unambiguous set expressions ("Please bring me the bill"). Ironically, because there is no real 'generative' production, the accuracy can be quite high, as we are simply copying the sound or symbol-image literally.

Category 2 is our ‘beginner’ stage: Small chunks of lexical terms start to have some meaning in highly constrained contexts; we can assemble extremely rudimentary sentences which express basic concepts, frequently using inter-language constructs:

“Man reading book” = A man is reading a book.

“Me ticket buy Liverpool” = I would like to buy a ticket to Liverpool.

Here the level of generation is essentially limited, with a small stockpile of names and actions that can be associated to communicate basic intentions or ideas. Nevertheless, with creative augmentation tactics to communication (i.e. Miming, gestures, use of mobile phone, etc.) a learner can nevertheless achieve more (in a functional context) than perhaps they would give themselves credit for.

Category 3 is our ‘intermediate’ stage: Here, vocabulary acquisitions and knowledge of grammatical structures is beginning to reach a level where the internal model starts to become increasingly ‘self-propelling’. The learner can navigate between a wide variety of contexts, link word-concepts in increasingly complex networks, and begin to map the language grammar to an increasingly sophisticated internal language model. They have reached the stage where inquiry within the language itself is possible, such they do not need to keep ‘exiting’ the language with recourse back to L1 to try to further their comprehension. At earlier stages, if one adheres to the communicative methodology, that necessarily means that certain language constructs are entirely unavailable without language ‘exit’; e.g. one could not practically explain the distinction between ‘determinism’ and ‘free will’ to beginner learners via a 100% pure communicative method. They would need to be at a higher level. A grammar translation method may yield some results with beginners, though the value of such

an exercise would be highly debateable given their pressing requirements for more basic, frequently occurring, and useful language terms first.

Category 4 is characterised by fluent ‘automaticity’; the person has relatively effortless control over language, reflected in no sense of nervousness or ‘self-consciousness’ in regard to the actual production of language. They are able to generate language constructs on demand for more or less any conceivable real life communicative situation, and are able to listen, speak, read and write all material typically representative of the general public. Note that someone could be ‘category 4’ – as are just about all native speakers who have received an education – yet that does not mean to imply their language use is ‘perfect’; many people’s native language written skills, for example, are often far from ‘perfect’, yet they are sufficiently advanced as to be still be constitutive of category 4 , as they can generally avoid producing any ‘inter-language’.

7. Socialisation & scaffolding

Language is a social phenomenon, and as such, language learning does not occur in isolation. Indeed, there are famous cases of abandoned human infants, that nevertheless managed to survive and grow up, who having never been exposed to any human language during their formative years, entirely seem to lack the capacity to develop any language in later years.

When you have a social group, the collective knowledge of the whole with regard to language acquisition is greater than that of any individual; though each individual may have either forgotten or simply not know any number of individual features, it may be that their peers have the knowledge they are lacking, and likewise.

Often knowledge/memory is tentative and is just missing a 'prop' to trigger the associated memory, since memory and meaning are deeply interlinked. The peer can provide the necessary input to trigger the associated memory/knowledge of language features through their provision of language use in context.

For example:

A: How you feel?

B: I feel good. What you like?

A: I like to drive car.

B: Where you drive?

A: I like to drive to house.

B: House where?

A: House that I live in.

B: Your home?

A: Yes drive to my home.

Here, the dialogue between person A and B is providing the necessary context for them to generate more language/inter-language than they would likely be able to produce individually if left alone. This process could and often can be amplified if you have a good peer learning group/environment; students that are particularly strong at certain aspects of language can offer the necessary context and modelling to assist other students with their acquisition of language in use. This process is called *scaffolding*, an evocative term that provides a near perfect metaphor since scaffolding is an intermediate construction that is used to support the construction of the main building (e.g. the cathedral, the office block), before eventually being removed once the target building is complete. Here too, scaffolding is an intermediate language support network that is removed and unnecessary once an individual has become fluent in a language.

8. Mediating factors between ‘input’ to ‘output’

There are many dozens of factors that influence the language acquisition process. We can broadly divide these into three components; input related, filter related, and output related.

Input related factors relate to how initial potential language information is received. I say 'potential' because unless certain pre-requisite abilities/cognitive skills related to the internal language model have been sufficiently developed, such input may not yield 'comprehensible' input. These include factors such as being able to recognise the phonological features of the language (e.g. can you actually consciously *hear* the five different tones in Thai?), to recognising the written symbols as letters, to whether or not you can detect grammatical structures that could provide a 'scaffolding' for comprehension of unknown terms, to whether or not the inputs occur in some sort of framed or confined context. Language is often used in such a way that the meaning is only clear when considered in the context in which it occurs, or sometimes even the wider context. All of this material needs to be suitably 'networked' together into some type of meaningful internal model, so that you can relate language items against concepts and terms that have meaning to you.

Filters are the larger 'extrinsic' factors that have a significant influence on the development of your language model. Are you motivated and concentrating? Are you attempting to apply yourself so you learn? Are you receiving material in different forms to assist with making it memorable? Does it create some sort of emotional engagement? (If there is no emotional component, there is likely to be no motivation. A person cannot really be 'forced' to learn something. The brain's internal heuristics tend to aggressively 'prune' and discard information/memories it considers as useless – why bother retaining memories if you are likely never to need them again?).

Output factors relate to the actual production and feedback processes involved in your language model. As soon as you have to speak (high time pressure) or write (lesser time pressure), there is a requirement that all of the information contained in the model is then actually *applied* to create some generative output. This is where the language ‘data’ gets ‘routed’ through the ‘network’ in order to yield an output. Certain ‘routes’, which routinely get used, may be relatively stronger, so output of these types of language structure are comparatively easier; developing our computer network analogy further, we can say this is rather like the routing of network IP packets over the TCP/IP protocol that underpins the internet, whereby two computers will gradually find the most efficient network ‘path’ to send their data packets, which normally means in practice that a file download rate should gradually increase in speed before stabilising at a relatively constant rate. Here the ‘packets’ are words or small groups of words and particular sentence structures, which with repeated use become established as a familiar ‘template’ from which to build an infinite variety of new, unique sentences.

Finally, rather like listening or reading has a direct physical action as a prerequisite, here the output depends on physical motor actions that either control the vocal cords to produce the correct sounds in order, or on the hands to scribe the correct shapes or type appropriate symbols into a computer in the target language.

9. Order of Acquisition

Despite the apparent gross differences between languages, generally speaking, there is a fairly prototypical sequence which learners progress through when learning a language. I have inserted a 'Stage 0', as though we take the token 'car' to refer to a single 'unit', in reality it is itself an aggregate of 'c', 'a', and 'r', three sounds that though intrinsically meaningless in themselves nevertheless acquire meaning when considered as structural components within an alphabet. The stages below are detailed with reference primarily to English, but it is possible in principle to abstract the stages to other languages through use of appropriate equivalent language specific structures:

Stage 0 – Individual 'atoms' (i.e. The individual characters of the alphabet or symbol system). The learner begins to associate basic patterns for aggregating these into words.

Stage 1 – Individual units (i.e. words or root lexical items). Often naming words, i.e. nouns. Stage 0 & 1 are effectively one stage, as a learner can often typically work on both simultaneously in lockstep; even young children can understand that a 'name', 'token' or 'word' consists of several sounds represented by letters, all of which combine to create an individual semantic link or unit (i.e. to mean or refer to something).

Stage 2 – By stage 2 the learner starts to have an awareness of the different functional properties of the units and how they can be contained combinatorially in order to satisfy the conventions established by the language's grammar. In English, the basic structure is SVO (like Thai); other languages may reverse or combine the elements in an entirely different order.

Stage 3 – Learners start to associate the different patterns required in order to satisfy the relationship between questions and imperatives. E.g. In English, they start to learn to ‘WH’ front questions, so instead of statements like ‘you live where?’ they learn to re-order into ‘Where you live?’.

Stage 4 – Learners start to develop a more complete syntactical knowledge of language structures, and start to appropriately re-order components according to the transformational rules present in the grammar. E.g. In English, they start to acquire an understanding of the appropriate markers corresponding to states of being or possession/ownership, such that “Is you tired?” might transform into “Have you tired?”. The complex and messy verb ‘to be’ coupled with the auxiliaries ‘do’ and ‘have’ start to become re-ordered and subject to appropriate transformations and substitutions. ‘Have are you going?’, ‘Where you are going?’, ‘Where are you going?’

Stage 5 – By this stage learners can start to re-arrange the full apparatus into nearly correct sentences. The ‘do’ auxiliary combines with ‘WH’ fronting so we get nearly correct questions such as ‘What do you have?’ mixed in with correct utterances.

Stage 6 – At stage 6 the negative form appears more frequently, though sometimes incorrectly ‘fronted’, and interrogatives related to ‘possibility’ emerge more frequently, such as ‘can’. ‘Can you visit cinema with me?’

Stage 7 – Full clauses and simpler sentences should now be possible to build with some accuracy. From here it is simply a matter of practice to develop the construction and combination of clauses with increasing sophistication and nuance.

10. Correcting inter-language mistakes

We have discussed above how inter-language is the production of incorrect linguistic forms that nevertheless exhibit meaningful features of the target language, and that this inter-language state is where a learner will spend the vast majority of their time before perhaps, or hopefully, reaching mastery.

Inter-language may be due to insufficient knowledge, or it may be related to insufficient capacity to *apply* the knowledge:

"I can hear in my head how I should sound when I talk, but it never comes out that way" (Gass & Selinker)

The solution is clearly more appropriate structured practice with suitable remedial input from a tutor where necessary. It is important to isolate whether the learner is producing errors or mistakes. If they are producing errors, they need more tuition, i.e. appropriate comprehensible input so they get suitable 'intake' which hopefully then causes remodelling associated with their 'integration'. If they are producing *mistakes*, then the solution is suitable remedial input from the teacher, which may involve deliberate correction to draw their conscious awareness to their mistake (a rather brutal method that can be discouraging and demotivating), or better, by recasting their language so they are given an opportunity to self-correct. If the extent of the *error* is not too large, recasting may work here too, provided the teacher or environment provides sufficient modelling or scaffolding for them to inductively build the new language knowledge 'on the fly'.

11. Tactics for correcting inter-language mistakes

a) Recasting

1 – “I had great time cycling. It was funny.”

2 – “Yes, we did have a great time cycling. It was fun.”

1 – “Very funny.”

2 – “Well, when you fell off the bicycle into those muddy leaves that was funny. The riding in general was fun. It was fun day out cycling.”

1 – “Yes, was a lot of fun”.

This example is an imaginary scenario of mine involving *recasting*, but based on a real-world inter-language mistake involving the use of the noun *fun* that I have observed among Thai speakers of English. ‘Fun’ can mean either something that was enjoyable, or something that caused amusement. In its adjectival form, only the second of these two meanings carries over. This is just one of the multitudes of vagaries of the English language. Here speaker 2 is trying to correct 1’s mistake by repeated recasting rather than explicitly informing them of their mistake, especially given that their mistake is actually the result of consistent and appropriate inferences from English language ‘rules’, so they are actually thinking and using language well. Since English grammar is a set of conventions rather than rules, this is why there are countless exceptions...

Recasting can be challenging as it is of no value simply to keep repeating the same ‘recast’ back if on each occasion the learner does not recognise which particular cue you are trying to correct. Instead, approaching the problem obliquely by ‘reframing’ the incorrect utterance may be enough to generate the necessarily self-correcting process. Since each learner’s internal model is unique, different approaches may need to be attempted to slowly focus in on the mistake. Here by involving humour, the teacher is attempting to provide a corrective cue in a manner that is won’t discourage the learner.

Original Recasting Attempt	Recasting – Alternative Attempt
<p>Alex: I ain’t got no pen Teacher: I don’t have a pen Alex: Ain’t you got a pen? Teacher: Don’t you have a pen? Alex: No. I ain’t got no pen. Teacher: They don’t have a pen. Alex: A’int nobody got no pen?</p>	<p>Alex: I ain’t got no pen Teacher: I don’t have a pen Alex: Ain’t you got a pen? Teacher: Did you have a pen this morning? Alex: Yes Teacher: So you <i>did</i> have a pen, but now you don’t have a pen? Alex: Yes. I don’t got a pen. Teacher: Here, borrow mine.</p> <p>{hands pen over}.</p> <p>Now <i>I</i> don’t have a pen... {class laughs}</p>

Another example. In this example, instead of offering ‘recasted fragments’ that simply lead to more misunderstanding, the teacher disguises their remodelling of the students mistake by framing it within a question thrown back at the student, giving a much more positive feel to the correction.

Original Recasting Attempt	Recasting – Alternative Attempt
<p>Anna: When the house built...</p> <p>Teacher: When was...</p> <p>Anna: When was?</p> <p>Teacher: When was. When was.</p> <p>Anna: When was what?.</p> <p>Teacher: The house built.</p> <p>Anna: That’s what I ask. When the house built?</p>	<p>Anna: When the house built...</p> <p>Teacher: OK. So you are asking when the house was built?</p> <p>Anna: Yes. When the house built?</p> <p>Teacher: When do you think it was built? So try and guess... When was the house built?</p> <p>Anna: Maybe 1960?</p> <p>Teacher: Close. So can you ask me the question again and I can tell you the answer?</p> <p>Anna: When was the house built?</p> <p>Teacher: Excellent. Well, the house was built in the 1950s.</p>

Some example errors/mistakes and possible resolutions. Errors may occasionally require a direct, explicit approach; tone and manner in which it is delivered can make a difference. If it is a new language construct the explicit correct should hopefully not seem ‘threatening’, particularly if we are not using the inappropriate teaching strategy of trying to introduce too much new material all in one go. Here the resolutions are oblique and indirect prompts that remodel the error into correct language..

Error/mistake	Resolution
1. He go to home.	<ul style="list-style-type: none"> Where is he going? Is he going home?
2. Where is book?	<ul style="list-style-type: none"> The book is on <i>the</i> table.
3. He came here since yesterday.	<ul style="list-style-type: none"> He came here since yesterday.
4. What do you like to do today?	<ul style="list-style-type: none"> I’m not sure. What <i>would you</i> like to do today?
5. Tell me what you doing?	<ul style="list-style-type: none"> I am studying for driving test. What are you doing?
6. Do you still in Saudi?	<ul style="list-style-type: none"> Yes, I still live in Saudi. Do you still live in Bangkok?
7. Where are you exercise?	<ul style="list-style-type: none"> I am going to exercise in the gym tonight. Are you going to exercise this evening?
8. What you favourite desert?	<ul style="list-style-type: none"> Is your favourite desert ice-cream? I love chocolate and mint. What is your favourite?

b) Clarifying questions

Student: I was rode my bike.

Teacher: You were riding your bike?

Student: Yes, riding bike.

Teacher: I was riding my bike too. Where did you ride?

Student: I ride around woods.

Teacher: I rode around my park too.

Here the teacher is presenting subtle corrections by presenting them in the form of ‘clarifying’ questions that seek to either obtain more information or seek confirmation about statements the student has made. In so doing they can provide the correct modelling for the language feature in a manner that is not discouraging to the student since they’re always presenting it in the form of seeking additional input rather than as an assertive explicit correction. The corrections they offer instead are implicit; the student can then learn to make amendments inductively.

12. A commentary on correcting in general

“Even if correcting the student’s every error were pedagogically feasible, socially acceptable, not demoralising to learners, and did not lead to their undue dependence upon the teacher, it still would not be psychologically sound practice. Pervasive correction ignores such important psychological limitations as memory capacity and attention span. Negative feedback, therefore, has to be judicious to be effective. However, selectivity is not sufficient in and of itself.” (Larsen-Freeman, 2003).

Larsen-Freeman highlight the unavoidable fragmentary nature of human memory and learning. Human memory is not like a computer, where there is a perfect storage of material provided it is input into in a correct format. Instead, even when the ‘comprehensible input’ is provided in the correct ‘format’, the retention will necessarily be imperfect and inputs will need repeated exposure to be consolidated. On the other hand, the more diffuse nature of human memory means that often a learner will actually know more than they are aware of or can credit themselves for; just because someone cannot explicitly recall something does not *automatically* mean they either do not know or have forgotten. Given what we can describe as the ‘emergent’ properties of language acquisition (and indeed memory and consciousness in general), some language features may simply be ‘dormant’ but nevertheless remodelling is occurring below the [self-]conscious threshold of the learner. What evidence do we have that suggests the presence of this phenomena?

The fact that someone, apparently not progressing at all, can suddenly make great progress. Human learning (whether learning language or how to play tennis) very rarely corresponds to a linear pattern. Instead it is bitty, fissiparous, and dynamic; sometimes it can superficially appear to stall, then suddenly a huge rate of progress can occur. Given this backdrop, the provisos in Larsen-Freeman are apt.

Above all, we have to exercise caution with correction, ensuring that it is of the right *quality* as well as appropriate in *volume* (i.e. amount). It is a balancing act. Too little correction and we are not providing sufficient feedback for students to improve. This is characterised by the common criticism of providing insufficient rigour. Conversely, too much rigour by employing excessive correction is demoralising. It tends to lead to a brutal winner-takes-all environment, which will suit only a small minority of learners who do indeed thrive, but most of whom will be left languishing discouraged and annoyed. So strictness must be balanced against inclusiveness when correcting.

The psychological state of the learner is critical for learning. If the language classroom becomes intimidating and unpleasant rather than enriching and fun, the balance is off; instead of excellent results you'll end up with dejection and disgust, a state entirely counterproductive to successful language acquisition.

13. The learning paradigm

Following Krashen, we can describe three general stages in this learning process of language acquisition:

1. *Comprehensible input*
2. *Reinforcement & Practice*
3. *Functionality*

In order to learn a language, a large amount of **input** must be provided; but as well as quantity, it must be at the appropriate level. If it is entirely incomprehensible, there are no language features that are recognisable, so the learner will simply hear white noise. If it is '**comprehensible**', i.e. $N+1$ just above the learner's current level, then appropriate scaffolding from peers, together with the teaching input and remedial correction will allow the learner to iteratively grow their language performance. This process occurs over a long period of reinforcement and practice, through controlled exposure and use of language features. Eventually, once the features have been mastered, they become a fully '**functional**' aspect of the learner's language. This process continues, in an iterative manner, until fluency has been reached. Even then, continual improvement is possible; native language speakers who choose to proactively study and read widely over their lifetime will gradually continue to add to their vocabulary and further refine their language capability.

Task 4 – Writing Performance Outcomes

- ▶ A set of slides with several hypothetical or potential performance outcomes for each activity based upon the example materials for a lesson sequence or series of lessons

1. Photograph

Three example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to look at the photograph and write six answers in a gap-fill sheet that has six specific questions regarding features in the photograph, with 75% accuracy.• Students will be able to look at the photograph and discuss in pairs for 2 minutes what they think the woman is thinking about. They will speak with around 50% accuracy: speaking fluency and communicative intention rather than grammatical precision is of target here.• Students will be able to look at the photograph and write down three basic sentences which describe what the woman looks like, where she appears to be, and whether they think she looks happy or sad. Students will write these sentences with 85% accuracy.	<ul style="list-style-type: none">• Teacher gathers gap-fill sheet and scores/marks according to the assessment rubric.• Teacher observes and monitors interactions. Peer based assessment in terms of continuing the dialogue.• Teacher collects/marks the three written sentences and scores/marks according to the assessment rubric.

2. Listening: Describing People

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to listen to the audio clip and tick the appropriate boxes to indicate which descriptive terms were mentioned in the audio with 80% accuracy.• Students will be able to listen to the audio clip and tick which boxes they think might describe the person who was talking. They will then discuss their answers with their partner for 3 minutes with 60% accuracy.	<ul style="list-style-type: none">• Teacher gathers tick-box sheet and scores/marks according to the assessment rubric.• Teacher observes and monitors interactions. Peer based assessment in terms of continuing the dialogue.

3. Listen & Describe 1

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to listen to the audio clip and write four down four items that the woman in the photograph has lost during the last month. They will be able to do this with 85% accuracy.• Students will be able to listen to the audio clip and discuss in a small group why the woman described is not the same woman as in the photograph. The group will then write down as many reasons/evidence (bullet points) as they can on the worksheet. Accuracy is based on total number of points identified from the list; 50% is realistic.	<ul style="list-style-type: none">• Teacher gathers gap-fill sheet and scores/marks according to the assessment rubric.• Teacher observes and monitors interactions. Worksheets get exchanged between groups and students will peer-mark each other's work based on teacher's descriptions.

4. Listen & Describe 2

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to listen to the audio clip and speak aloud with at least two answers as to what the pictured boy thinks of his mother (previous slide);. Students will speak with 60% accuracy.• Students will be able to listen to the audio clip and describe how the boy spent his day by recording themselves speaking for at least 30 seconds into the digital audio recorder/PC with on-screen questions to act as prompts, with 50% accuracy.	<ul style="list-style-type: none">• Teacher listens to the students speaking utterances and records appropriate assessment notes in a rubric/sheet.• Teacher gathers/listens the audio clips from the classroom network storage and scores against a rubric.

5. Describe yourself

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to write down five sentences which describe facts about themselves and their day-to-day life with 75% accuracy.• Students will be able to work in pairs to answer a set of 20 yes/no questions on the worksheet with descriptive features about themselves (e.g. “I am a very tall person”; “I like to go running”). One student will read aloud and the other student will listen and circle the sheet with their answer. They will then swap roles. Students will read aloud with 80% accuracy and answer with 90% accuracy.	<ul style="list-style-type: none">• Teacher listens to the students speaking utterances and records appropriate assessment notes in a rubric/sheet.• Teacher monitors the speaking activity. Students peer review each other’s answers and scaffold their speaking where necessary.

6. Conversation

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to read the dialogue script and speak aloud with their characters lines with 70% accuracy (swap roles and repeat).• Students will be able to listen to the audio clip of the conversation and mark exactly where in the script the conversation has got to each time with a number as the teacher presses the stop button on the audio clip with 85% accuracy.	<ul style="list-style-type: none">• Teacher listens to the students speaking utterances and records appropriate assessment notes in a rubric/sheet.• Students swap scripts and peer-mark. Teacher elicits the answers from the class as a whole class informal marking activity.

7. INFO GAP A & B

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to Q&A in pairs and write their answers in their info gap sheet from the others corresponding answers with 95% accuracy.• Students will be able to listen to the audio clip for their partner's worksheet with headphones (their partner cannot hear it). They will then speak aloud and the other student will write their answers to complete their info gap worksheet. The roles will then reverse with the second audio clip played. The students will fill in in their info gap sheet with 70% accuracy. (The two info gap sheets do not 'complete' each other as a pair in this version).	<ul style="list-style-type: none">• Teacher listens to the students speaking utterances and records appropriate assessment notes in a rubric/sheet. Teacher gathers info gap worksheets and assesses completeness.• Students swap scripts and peer-mark. Teacher elicits the answers from the class as a whole class informal marking activity.

8. Focus on Form– Fill in Grammar

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to read a descriptive paragraph of text and write the appropriate answers in the 13 blank spaces on the worksheet with 65% accuracy.• Students will be able to watch a video (played twice in total) and will write the appropriate answers to the descriptive questions in the 13 blank spaces on the worksheet with 75% accuracy.	<ul style="list-style-type: none">• Teacher gathers worksheets and marks according to the set answers.• Students swap scripts and peer-mark. Teacher elicits the answers from the class as a whole class informal marking activity.

9. Focus on Form – Vocabulary

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to read the 10 sentences and write the correct descriptive word from the seven possible options in the space provided with 45% accuracy.• Students will be able to work in pairs to match the correct word (sliced off into individual rectangles) with the corresponding sentence with 60% accuracy.	<ul style="list-style-type: none">• Teacher gathers worksheets and marks according to the set answers.• The paired students will speak aloud with their answer to one question to the whole class which will then check their answer as an informal class discussion/marking activity.

10. This Is My Family

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to read the 6 sentences and write the correct descriptive word from the seven possible options in the space provided with 45% accuracy.• Students will be able to work in pairs to match the correct word (sliced off into individual rectangles) with the matching sentence with 60% accuracy.	<ul style="list-style-type: none">• Teacher gathers worksheets and marks according to the set answers.• The paired students will speak aloud with their answer to one question to the whole class which will then check their answer as an informal class discussion/marking activity.

11. What's Your Family Like?

Two example potential performance outcomes

Students do	Assessment
<ul style="list-style-type: none">• Students will be able to write five sentences of their own creation which describe their family with 60% accuracy.• Students will be able to write about 2-3 other family members and use the seven previous vocabulary items to make basic sentences about them with 75% accuracy.• Students will be able to work in groups of three to select 2-3 characteristics of their family from a pot full of set sentences which they will then role-play to the other two members of the group (“e.g. Granddad is always grumpy”; “Mother is always in a hurry”; “Dad is always late for work”). Students will role-play with 90% accuracy.	<ul style="list-style-type: none">• Teacher gathers worksheets and marks according to the set answers.• Teacher gathers worksheets and marks according to the set answers.• Teacher observes the role-play groups and makes notes about student speaking-listening on each group in turn.

Task 4 – Assessment Based Language Education (ABLE)

- ▶ A set of slides to demonstrating my understanding of Assessment Based Language Education.
 1. The traditional model...
 2. ...Vs Assessment Based Language Education
 3. Why ABLE?
 4. Writing Performance Outcomes
 5. Example Planned Performance Outcomes

1. The traditional model...

The traditional teaching model is one based upon defining a set of *goals*, devising a set of teaching *actions* and supporting materials derived from these goals, and then finally arriving at the learning *outcomes*. This ‘rational-linear’ model (i.e. a ‘rational’ pre-determination of goals followed through with the necessary steps in a ‘linear’ manner) is actually orientated around the *teacher* rather than the students. It is essentially ‘practice-centric’. You could also describe it as a ‘top-down’ model. The traditional model looks something like this:

1. Goals/Objectives: *How* shall I teach it?

Teacher-centric: The teacher starts from the basis of what they *believe* students *should* know (objectives), and how they best believe that this can actually be *taught*.

Strong determining factors:

- ▶ **Personal beliefs & principles**
- ▶ **Personal preferences toward teaching style**
- ▶ **Personal analysis of suitable material**
- ▶ **Personal preferences toward student activities**

2. Actions/Activities: *What shall we teach?*

Syllabus-centric: Having decided upon a set of learning goals for the students in accordance with their teaching principles and beliefs, the teacher then scrutinises the syllabus and any set texts, selects appropriate materials, and on this basis decides exactly what they will teach.

Strong determining factors:

- ▶ Conveniently available materials
- ▶ Contents of the official curriculum/syllabus
- ▶ What activities do I like to do with students in the classroom?

3. Outcomes/Tests: *What will the students be able to do and what tests shall we set to assess this?*

Testing-centric: Having decided upon the how and what to teacher, and then having selected and designed materials for this based upon the contents of the syllabus/curriculum, the final stage is to gather evidence of student learning of this material by designing appropriate tests.

Strong determining factors:

- ▶ Amenability of material to be testable
- ▶ Content of the syllabus
- ▶ What types/styles of test do I like to administer to students in the classroom?
- ▶ Ease of gathering data/testing convenience

2. ...Vs Assessment Based Language Education

So what is Assessment Based Language Education (ABLE)? The name gives a clue. Essentially it **reverses** the traditional model described above by *starting* from the learning outcomes and our means for **assessing** these.

By starting from the *outcomes* rather than from how/what we would like to teach, the entire process becomes much more learner-orientated. So the new model looks like this:

1. Outcomes/Assessment

Learner-centric: What do we want the student to be able to actually *do*?

On the basis of what we want the student to *do*, exactly *how* will we test/assess for this, in order to determine whether they can or cannot *do* whatever it is we intend for them to be able to do?

Our intended performance outcome for the student determines our testing procedure – *not* the other way round!

2. Actions/Activities: *What actions do we need to take, and what activities do we need to do, in order to reach our learning outcomes?*

Outcome-centric: In ABLE, having clearly identified what learning (i.e. performance) outcomes we want our students to actually accomplish, we can therefore now decide upon what teaching actions (i.e. ‘inputs’) we need to provide, and select/design our materials in accordance with these actions. Instead of selecting and devising material simply in accordance with a syllabus, whatever is convenient or to hand, or whatever happens to suit our particular teaching style/approach/philosophy, instead we make the decision on the basis as to what material will help enable the achievement of the outcomes by our students.

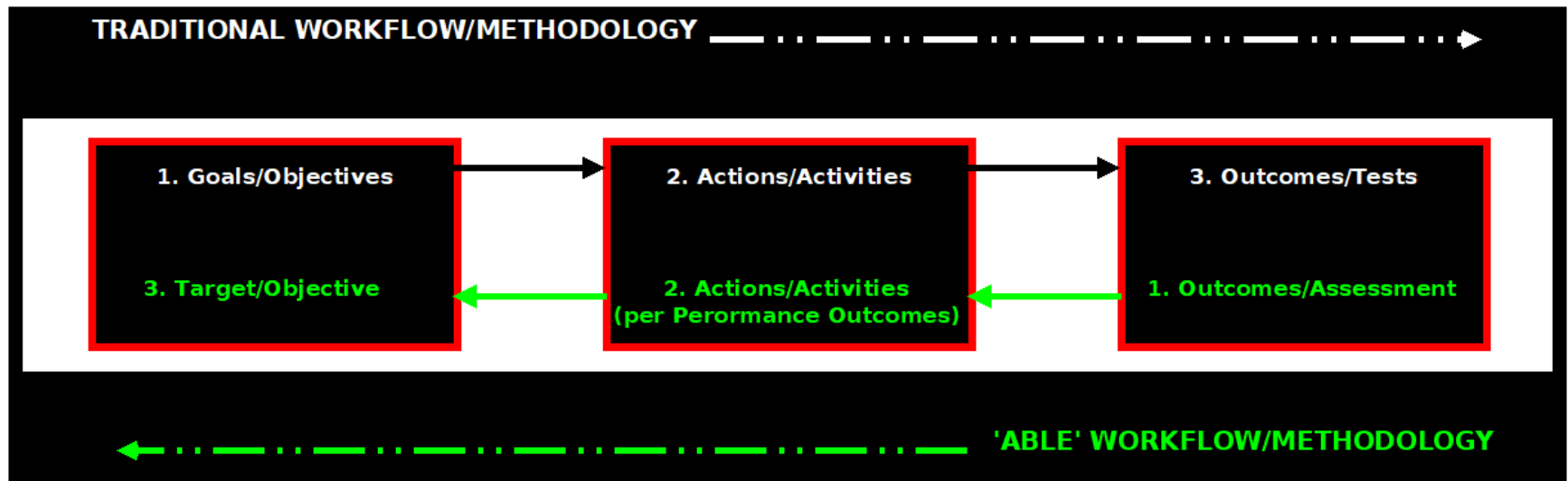
Our actions/activities are designed/selected in accordance with fulfilling our intended performance outcomes for students. What they actually *need* in order to produce the outcomes we have identified.

3. Target/objectives: *On the basis of the content/materials we need to provide in order to accomplish our performance outcome, how can we best teach this?*

Objective-centric: From steps 1 and 2 we have identified what we want our students to do, and on the basis of that what content (i.e. materials and activities) we need to deliver to them in order to help enable it. The final step in this equation, working backwards, is then to identify *how* we shall teach it.

This should actually not be a particularly abstract step. The better 1.) and 2.) have been planned, the simpler 3.) should be, in principle; with a very clear set of performance outcomes, and material specifically selected to enable that outcome, the actual process involved in teaching that should be to a large degree 'revealed'. Our 'how' becomes a matter of mapping student actions from the content of 2.) to the 'output' of 3.). This 'learner' orientated process will have revealed the precise actions needed to be taken. The teacher is then able to facilitate this process by deploying whatever techniques are actually required in the particular classroom situation to assist students through this learning process.

My diagram illustrating ABLE vs. the traditional planning methodology:



3. Why ABLE?

So ABLE reverses the traditional model. So why do this? What advantages does it give us? By reversing the model, we have made the education process task-specific. Rather than start with a somewhat vague or distant set of goals (e.g. I want to teach these students how to read English to an intermediate level, I want to teach them about Renaissance humanism in the Italian city-states) with the result that there is no clear process as to exactly how this will be achieved, instead, by starting from the learning outcomes, it forces us to be highly specific about our objectives right from the beginning of the planning process. By very specifically identifying what we want our students to *do*, it forces us to then consider precisely what content and activities (i.e. input) will be required to give them the desired skills/knowledge. Having then decided on the material, only then do we finally consider how to teach it; this process having been made all the easier since we now know *what* it is we are going to teach.

ABLE might seem unintuitive, but by planning 'backwards' it actually forces a highly specific and concrete method of planning. Conversely, starting forwards in the conventional rational-linear model could end up with you missing the actual end-target (the students actually being able to do or demonstrate the knowledge in the particular ways you desired), and worse, by the time you realise this, it is likely too late in the process (i.e. Not enough term/teaching time left to address all the problems). ABLE is learner-centric, the traditional model is teacher-centric; but teaching is ultimately about what the learners end up knowing/being able to do, not about which particular means were used to get there.

4. Writing Performance Outcomes

The critical heart of ABLE is the learning/performance outcomes.

Given that these should be as specific and concrete as possible, how do we go about writing them? One simple method is to consider them in terms of a set of ‘action’ verbs in English. So, for TESOL, we might break them down into three categories broadly stated as ‘saying’, ‘writing’, or otherwise ‘showing’ via some manner. Other words and categories could be used; the important point is that the words that are used in the performance outcome relate to some highly specific action that the student will be ‘doing’ that we can ‘see’; they will be producing a specific piece of evidence that we can by some means assess. So:

Say	Write	Show
Speak	Spell	Act
Tell	Complete	Role-play
Describe	Draw	Charade
Converse	Trace	Point
Repeat	Colour	Mime
Pronounce	Label	Gesture
Reply	Fill in	Match
Sing	Construct	Mouth

We can then further refine the writing of performance outcomes by considering it from the perspective of four dimensions:

► **What will they do?**

From the action verbs table above; label, mime, role-play, describe, reply, etc...)

► **How will they do it?**

Individually? In pairs? In small groups? As a class mingle?

► **With whom will they do it?**

Fellow students? With the teacher? Group to group?

► **To what degree of accuracy?**

A rough estimate based on previous experience and realistic expectations based on task difficulty; e.g. 70%, 55%.

5. Example Planned Performance Outcomes

Here are ***three*** examples of mine that bring this all together:

- ▶ *1: "Students will be able to **individually spell** vocabulary words encountered earlier in the lesson from audio prompts with **65% accuracy**"*
- ▶ *2: "Students will be able to **role-play in pairs** the functional English on buying & selling common food items in the market with **50% accuracy**"*
- ▶ *3: "Students will be able to **pronounce** the words written on the whiteboard relating to transport as a **whole class choral drill** with **90% accuracy**"*

Task 5 – Testing Skills Input

(Group jigsaw task)

This jigsaw activity has presented a bit of dilemma to me as each groups work was radically different in terms of formatting/presentation and approach, and there are of course variances in quality and quantity of work. In an ideal world I would like to have been able to rewrite/completely rebuild every set of group slides to bring it to the same standard and present it all in a consistent manner here. In particular, some groups would benefit from expanded and improved performance outcomes, but this would then involve restructuring some of the tasks and activities they have devised, or in many cases even require constructing entirely new ones from scratch.

However, such a task is completely infeasible in the time constraints I have (it would probably take me a further 20 hours of work or more to completely build a new set of slides from scratch, drawing on ideas that each group has presented to get anywhere near the standard I would like) and I am conscious of the quantity of slides in this assignment in any case. Even reformatting every single slide into the same format as this assignment would take a considerable time. So, instead, my compromise approach is this:

1. Present a short critique/set of notes for of each group's work as a summary of each groups work. These critiques can be found on the following slides:
 - [Testing Listening Critique/Notes \(Slide 100\)](#)
 - [Testing Speaking Critique/Notes \(Slide 145\)](#)
 - [Testing Reading \(My group\) Critique/Notes \(Slide 154\)](#)
 - [Testing Writing Critique/Notes \(Slide 185\)](#)
2. Create my slides for grammar testing according to the model I would prefer, within the realistic time constraints I have (it could of course be expanded further). This section starts on slide 186:
 - [Testing Grammar \(Slide 186\)](#) .

Task 5.1 – Testing Listening

Slides by:

Alvin Uy

Ismail Vorajee

Luke Di-Maria

Cody Naidoo

Andrea Lobrigas

Testing Listening

Alvin Uy
Ismail Vorajee
Luke Di-Maria
Cody Naidoo
Andrea Lobrigas

Weather

- Formative Outcomes
- Performance Outcomes
- Context and Modeling
- Listening Activities
 - Alvin Uy - Matching names and photos also T/F Qs
 - Ismail Vorajee - Fill in the blanks
 - Luke Di-Maria - Multiple Choice
 - Cody Naidoo - Chinese Whispers
 - Andrea Lobrigas- Conversation Script

Chosen CEFR Level

A2 - Can understand enough provided speech is clearly and slowly articulated

B1 - Can understand the main points of clear standard speech on familiar matters regularly encountered in work, school etc

Formative Outcomes

Listen to audio/script and **make** notes.

Extract and write down info from a conversation audio.

Fill in the blank boxes with T/F.

Fill in blanks from listening to audio.

Listen and **complete** multiple choice Qs.

Participate in Chinese Whispers Run and Read game. *LF weather and days of the week.*

Work in pairs and read through the conversation.

Performance Outcomes

Student does	Teacher does	Assessment
Listens to the audio and makes notes	Provides photo and audio for warmer	Takes notes, records student on video
Matches names and photos	Models matching activity	Records the grades
Listens to the audio again and answers the T/F Qs	Models T/F Qs	Collects and records the grades
Fills the blanks with correct answer/s	Models fill in the blank activity	Records correct answers and score
Checks the notes and completes multiple choice Qs.	Models Multiple choice worksheets	Records how many Qs are completed correctly.
Participates Chinese whisper game Listens to the instruction Participates in the games Finds a groupmate Listens to a groupmate carefully Runs to the other groupmate Writes using LF on the board	Models group activity Models instruction Models board work	Watches, records student on video
Participates in the pair activity Writes a conversations Acts through conversation	Models pair activity Models role play	Watches, records student on video/ audio

Context and modeling (Warmer)

SCRIBBLE WEATHER

5 Day Forecast

Monday



Tuesday



Wednesday



Thursday



Friday



Script (will be done through audio)

Monday - cloudy to overcast with some showers spread throughout the day.
Highest temperature of 14 degrees and lowest temperature of 12 degrees.

Tuesday - mainly dry and cloudy, with a chance of some drizzle late afternoon.
Highest temperature of 15 degrees and lowest temperature of 9 degrees.

Wednesday - partly cloudy, dry with chance of showers. Highest temperature of 10 degrees and lowest temperature of 7 degrees.

Thursday - cloudy to overcast with some showers spread through the day. Highest temperature of 13 degrees and lowest temperature of 8 degrees.

Friday - showers with some dry patches late afternoon. Coldest day of the week.
Highest temperature of 7 degrees and lowest temperature of 4 degrees.

Matching names and photos (Alvin Uy)

Monday



Tuesday



Wednesday



Thursday



Friday



- TASK: Fill in the blanks. Which one is,
 - Cloudy to overcast with some showers spread throughout the day. Temperature: Highs - 14C, Lows - 12C.
 - Mainly dry and cloudy, chance of some drizzle late afternoon. Temperature: Highs - 15C, Lows - 9C.
 - Partly cloudy, dry with chance of showers. Temperature: Highs - 10C, Lows - 7C.
 - Cloudy to overcast with some showers spread throughout the day. Temperature: Highs - 13C, Lows - 8C.
 - Showers with some dry patches late afternoon. Temperature: Highs - 7C, Lows - 4C.

True or False (Alvin Uy)

Monday	T	F	Tuesday	T	F	Wednesday	T	F	Thursday	T	F	Friday	T	F
Cloudy to overcast with some showers spread throughout the day.			Mainly dry and cloudy, chance of some drizzle late afternoon.			Partly cloudy, dry with chance of showers			Hot, no one can survive.			Showers with some dry patches late afternoon		
Highest Temperature: 14C			Highest Temperature: 20C			Highest Temperature: 10C			Highest Temperature: 50C			Highest Temperature: 7C		
Lowest Temperature: 10C			Lowest Temperature: 9C			Lowest Temperature: 7C			Lowest Temperature: 40C			Lowest Temperature: 4C		

Weather blank fill - Ismail Vorajee

Monday



Tuesday



Wednesday



Thursday



Friday



--	--	--	--	--

Listens and answer - Multiple Choice (Luke Di-Maria)

1. What's the weather mainly like on Tuesday?

- a. dry
- b. wet
- c. hot
- d. windy

2. What is the lowest temperature on Thursday?

- a. 8 degrees
- b. 10 degrees
- c. 9 degrees
- d. 7 degrees

3. What is the coldest day of the week?

- a. Monday
- b. Thursday
- c. Friday
- d. Tuesday

4. What is the highest temperature on Wednesday?

- a. 6 degrees
- b. 10 degrees
- c. 9 degrees
- d. 7 degrees

Chinese Whispers - Runs and Writes using LF (Cody Naidoo)

LF - The weather is _____ on _____.

Monday

dry

Tuesday

wet

Wednesday

cloudy

Thursday

hot

Friday

cold



Conversation Script - Andrea Lobrigas

A: I really want to go to the mall this Wednesday.

B: That sounds like fun. What's the weather going to be like?

A: I heard it's going to be nice but cloudy.

B: Is it a perfect time to go shopping.

A: I believe so, it's 10 C, so it won't be so hot.

B: Good. I hope it doesn't rain.

A: I know. I really want to go to the mall to buy stuff for the party.

B: But you know that California weather is really unpredictable.

A: You're right. One minute it's hot, and then the next minute it's cold.

B: I really wish the weather would just stay the same.

A: I do too. That way we can have our activities planned ahead of time.

B: Yeah, that would make things a lot easier.

Task 5.1 – Listening Group Critique/Notes

- ▶ The slides would benefit from having the summative as well as formative outcomes listed.
- ▶ Some of the performance outcomes are too vague to be useful/assessable. “Listen to the audio clip and make notes”: make notes on *what* exactly? There is insufficient direction to students listen intensively for a purpose. More precision and expansion of the performance outcomes is required.
- ▶ The weather task seems to be assessing students particular interpretations of the weather symbols as much as it is assessing their actual listening skills, which is the intended goal. That being said, it could be made workable with adjustments, and has merit as an approach.
- ▶ In general, the slides would benefit from more clarification/expansion.

Task 5.2 – Testing Speaking

Slides by:

Nigel Mullally

Lee McGill

Malcolm O'Donnell

Curtis Haran

Speaking Test Group



Nigel Mullally, Lee McGill, Malcolm O'Donnell and
Curtis Haran

Secondary Speaking Assessment

CEFR B1:

Students can exploit a wide range of simple language to deal with most situations.

Students can say short/simple sentences.

Students can give short/simple answers to questions.

CEFR B2:

Students can interact with a degree of fluency and spontaneity that characterizes regular interactions.

Summative Outcomes

Students will be able to:

- Ask and answer questions about travel plans for the weekend.
- Talk about activities situated in appropriate locations.
- Ask & answer questions about travel companions

Formative Outcomes

- **Greet** each other, **say** and **spell** names. (10)
- Observe T, **elicit** words based on T's context setting. (10)
- **Ask & answer** questions about plans for the weekend. (15)
- **Perform** dialogue in pairs and groups
- Listen to audio **mark** answers, **answer** Qs, **say** sentences. (15)
- In groups Students watch a video take **notes**, **write** own dialogue then read out new dialogue for video (15)
- S/S **ask** and **answer** (Where and What) questions in context setting for activity(20)
- In pairs, **ask & answer** questions about plans for the weekend, **correcting** any grammar errors if necessary. (25)
- **Write** their own dialogue, **perform** dialogue (15)

Assessment Areas within the Lesson

1. Warmer: **Greet** others **Ask** and **Ans** Q, **spelling** aloud.
2. Context/Model: **notes** and **writes** TL, **underlines** KWICS and **says** the dialogue.
3. Controlled Practice: group **writes** posters, **writes** to WB, **speaks** TL-TS, ST, **says** dialogue in pairs.
4. Vocab Ex: Listen and **hit** flashcards with a toy hammer.
5. Listen, Mark, Answer and Say. **Ask** and **Ans** Qs, **Say** dialogue
6. Audio creation: **Write** new dialogue and **read** out with video, S/S
7. Communicative Activity: **Asking** and **answering** questions, **grammar**, **structure** and **spelling** T/S, S/S
8. Paired Interview: **Ask** and **Ans** questions in pairs SS referring to flashcards.
9. Presentation: **Write** dialogue and **perform** dialogue

These 9 assessment areas, which have been further divided into at least 3 metrics have been additionally delineated by oral (blue) and other skill areas (green), and this should allow for adequate triangulation.



Assessment Criteria and Rubrics

B1	Grammar and Vocabulary	Discourse Management	Pronunciation	Interactive Communication
5	<ul style="list-style-type: none"> Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on familiar topics. 	<ul style="list-style-type: none"> Produces extended stretches of language despite some hesitation. Contributions are relevant despite some repetition. Uses a range of cohesive devices. 	<ul style="list-style-type: none"> Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly. 	<ul style="list-style-type: none"> Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.
4	<i>Performance shares features of Bands 3 and 5.</i>			
3	<ul style="list-style-type: none"> Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about familiar topics. 	<ul style="list-style-type: none"> Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, but there may be some repetition. Uses basic cohesive devices. 	<ul style="list-style-type: none"> Is mostly intelligible, and has some control of phonological features at both utterance and word levels. 	<ul style="list-style-type: none"> Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.
2	<i>Performance shares features of Bands 1 and 3.</i>			
1	<ul style="list-style-type: none"> Shows sufficient control of simple grammatical forms. Uses a limited range of appropriate vocabulary to talk about familiar topics. 	<ul style="list-style-type: none"> Produces responses which are characterised by short phrases and frequent hesitation. Repeats information or digresses from the topic. 	<ul style="list-style-type: none"> Is mostly intelligible, despite limited control of phonological features. 	<ul style="list-style-type: none"> Maintains simple exchanges, despite some difficulty. Requires prompting and support.
0	<i>Performance below Band 1.</i>			

Speaking Grade Sheet

Name	Grammar/Vocab	Pronunciation	Communication	Discourse
Ploy	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Performance Outcomes **Warmer**

Students Do	Teacher does	Assessment
<ul style="list-style-type: none"> • Listens to T • Responds and initiates • Asks & answers 	<ul style="list-style-type: none"> • TT models • TS>ST>SS 	T notes the general level of the class on the previously taught language: greetings, names, spelling.
Ask & answer in double lines	Models //”s and S switch	
Substitute <ul style="list-style-type: none"> • Alternative language in the greeting/ request- How’s it going. Can you spell it please? 	<ul style="list-style-type: none"> • T models alternative language for the greeting and request 	
	T refers to the student speaking rubric when making notes and	Note. This language was taught in a previous class and this is distributed

Performance Outcomes Modeling

Context &

Student does	Teacher does	Assessment
Looks, chats, repeats and makes notes	Models the context and the target language on the board and at the front of the class	<ul style="list-style-type: none"> Monitoring student notes Notes are made as a group on large posters
<ul style="list-style-type: none"> underlines/highlight key language features as shown by T Copies/draws T's example from the board 	<ul style="list-style-type: none"> Models the context Models the key lexis for S Highlights the grammatical, contextual and pronunciation features 	<ul style="list-style-type: none"> Takes photo of S poster T rates S using rubric and grades T audio/video record S speaking for later assessment using rubric
<ul style="list-style-type: none"> Participates in 	<ul style="list-style-type: none"> Sets up groups for 	<ul style="list-style-type: none"> T makes notes and

Performance Outcomes Practice

Controlled

Student does	Teacher does	Assessment
<ul style="list-style-type: none"> • Observe TT models • TS>ST>SS • Writes Tar Lang 	<ul style="list-style-type: none"> • Models TT of conversation • 2 lines at a time until all six are on the WB • TT>TS>ST>SS>WB x3 • T cleans the WB • Ss in groups rewrite the dialogue • T black random non key lexical/ cotextual features 	<ul style="list-style-type: none"> • SS in small groups write core dialogue on large poster • T makes notes of Ss retention and production ability of core language • T notes the ability if individual S to recreate the dialogue on the WB and in their notes
<ul style="list-style-type: none"> • Ask and answers Ss Q in double lines 	<ul style="list-style-type: none"> • T divides Ss in half • Sets up double lines • Models the S switch 	<ul style="list-style-type: none"> • Observes and makes notes of Ss performance of Tar

Performance Outcomes: Listen, Mark and Answer

Student does	Teacher does	Assessment
Ss respond to T, In pairs Ss listen to the audio for Qs, make notes, practice asking the Q's S/S	T models Qs, KWICs, gives instructions, elicits with ICQs plays audio T/S S/T <ul style="list-style-type: none">• Where is he/she going?• What is he/she going to do there?• Who is he/she going with?	Observes Ss accuracy, pronunciation, takes notes
Ss respond to T, in groups Ss listen to the audio for answers, discuss, practice asking and answering Q's answers SSS	T models Qs, KWICs, gives instructions, elicits with ICQs, plays audio T/S S/T <ul style="list-style-type: none">• He's/she's going to the _____• He's/she's going to	Observes Ss accuracy, pronunciation, takes notes

Performance outcomes

Video, Audio creation and role play P

Student does	Teacher does	Assessment
Ss watch a video with no Audio and take notes	Demonstrates activity by doing own Dialogue for video T/T <ul style="list-style-type: none">• Where are we going?• To the beach.	
Ss in groups write their own Dialogue S/S	Walks around assisting students if required.	Assesses and takes notes
Ss practice their dialogue in their groups S/S	Observes Ss	
Ss present their dialogue to the class S/S	Teacher turns on video	Takes notes on structure and grammar accuracy

Performance outcomes

Communicative Activity

Student does	Teacher does	Assessment
Ss Listen to instruction	Teacher explains the activity and gives out activity cards	Checks understanding
Ss take listen and take notes.	T/T demonstration... <ul style="list-style-type: none"> • Where are you going? • I'm going to the beach. • What are you going to do? • I'm going swimming 	Monitors Ss notes, records
Ss walk around the room Q&A and fill in the name of the Student that is going the beach etc	Teacher walks around the classroom to supervise the activity and check student understanding	Takes notes
Ss Q&A T/S S/S S/S After first teacher ? Students take over. Ling want to the beach	Teacher asks Where did Ling go?	Assesses accuracy and takes notes

Performance outcomes

Paired Interviews

Student does	Teacher does	Assessment
Observe teachers task instructions.	Demonstrates activity by asking volunteer from class. <ul style="list-style-type: none">• T.Where are you going? (shows flashcard)• S. I'm going to the beach.	Looks for understanding.
Ss in pairs Q + A regarding flashcards	Hands out flash cards,Walks around assisting students if required.	Takes notes on structure and grammar accuracy
Move and find a different pair	Observes	N/A
Ss Q + A, correct grammar if necessary	Observes	Takes notes on structure and grammar accuracy

Performance Outcomes

Presentation

Student does	Teacher does	Assessment
Respond to T, in pairs Ss discuss and write their own dialogue S/S	T gives instruction, models example, elicits with ICQs T/S S/T SSS	Observes and takes notes
Students practice their dialogue S/S	T observes	Observes and takes notes
Ss present presentation	T observes	Assess Ss speaking accuracy and pronunciation accuracy by end of 1st 20%

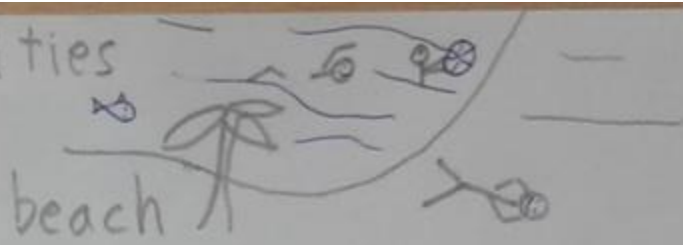
Warmer

- A. Hi. My name's !!!!!. What's your name?
B. My name's Ploy.
A. How do you spell that?
B. It's P-L-O-Y.
A. P-L-O-Y. Ploy. Nice to meet you.
B. Nice to meet you too.



Context & Modeling

Weekend Activities



this weekend

Mon	Tues	Wen	Thur	Fri	Sat	Sun
1	2	3	4	5	6	7
8	9	10	11	12	13	14

- A. Where are you going this weekend?
B. I'm going to the beach.
A. What are you going to do?
B. I'm going to have a picnic.
A. Who's going with you?
B. My sister and some friends.

Core Dialogue

- Ploy: Where are you going this weekend?
- Tom: I'm going to the beach.
- Ploy: What are you going to do?
- Tom: I'm going to have a picnic.
- Ploy: Who's going with you?
- Tom: My sister and some friends.

Lesson Sequence

- Conversation
- Ploy: Where are you going this weekend?
- Tom: I'm going to the beach.
- Ploy: What are you going to do?
- Tom: I'm going to have a picnic.
- Ploy: Who's going with you?
- Tom: My sister and some friends.

—Model to the **BB**

—**TT>T/S>S/T>S/S>BB**

—**Doubles-lines/Horse shoe**

Controlled Practice: Weekend Activities

- Ploy: Where are you going this weekend?
- Tom: I'm going to the beach.



Weekend Activities #2

- Ploy: What are you going to do?
- Tom: I'm going to have a picnic.



Weekend Activities #3

- Ploy: Who's going with you?
- Tom: My sister and some friends.



Vocab Expansion Keys

1. Beach	1. Father	1. Swimming
2. Gym	2. Sister	2. Hiking
3. Mall	3. Mother	3. A Coffee
4. Park	4. Brother	4. Exercise
5. Waterpark	5. Friends	5. Sports
6. Salon	6. Mother	6. Relax
7. Rollercoaster	7. Pet	7. Picnic
8. Swimming Pool	8. Grandfather	8. Have fun
9. Coffee Shop	9. Grandmother	9. Shopping

Vocab Expansion 1 - places to go



Vocab Expansion 2 - People

1.



3.



6.



8.



2..



4.



7.

5.



9.



Vocab Expansion 3 - Activities

1.



2.



3.



4.



5.



6.



8.



9.



7.



Listen, Mark, Answer and Say



John: Where are you going this weekend?

Peter: I'm going to the city.

John: What are you going to do there?

Peter: I'm going shopping

John: Who are you going with?

Peter: I'm going with my grandmother.

Q: Where's Peter going?



Q: What's he going to do there?



Q: Who's he going with?



Listen, Mark, Answer and Say



Rose: Where are you going this weekend?

Jane: I'm going to the pet shop.

Rose: What are you going to do there?

Jane: I'm going to buy a rabbit.

Rose: Who are you going with?

Jane: I'm going with my Brother.

Q: Where's Jane going?



Q: What's she going to do there?



Q: Who's she going with?



Video, Audio creation and role play



Students watch a video of a family going to the beach with no Audio.
Students then in groups create their own Dialogue and role play



group to group so they're learning and teaching.

Example of what Students could produce.

John: Dad where are we going?

Dad : We're going to the beach.

Lisa : I like the beach. How are we going there?

Mum: We're going by plane.

Lisa: Mum, can I have an Ice cream?

Mum: Yes. Anyone else like an ice cream?

Dad: No thank you.

John: No, thank you.

Lisa: Dad, where are you going?

Dad: I'm going for a swim. What are you doing?

Lisa: I'm going to make a sandcastle with mum.

Mum: Lisa lets build a castle

Lisa: No, I'm going to sleep now.

Communicative Activity



Beach



Swimming



Cinema



Avengers



Camping



Scout Camp



The Farm



See Grandma

Communicative Activity

Where are you going?



What are you doing?



Communicative Activity

Where are you going?



Ling



Wing



Ming



Bing

What are you doing?



Ling



Wing



Ming



Bing



Communicative Activity

Students will first receive a card showing where they're going and what they are doing.

Then they will get a second card, either Where are you going or What are you doing.

Then the students will walk around Q and A to find out who went where and who did what.

Finally when they have completed all 4 boxes they will speak with the Teacher.


Beach

Swimming

Where are you going?		What are you doing?	
			
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
			
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Teacher: Where is Ling going this weekend?

Ming: Ling is going to the beach.

Teacher: What will she do?

Bing: She's going swimming.

Paired Interviews (grammar activity)

1) Firstly, the pairs collect 2 place, 2 person and 2 activity flashcards and divide them between the pair.

2) Then the pairs ask each other 3 questions:-

- Where are you going?
- What are you going to do there?
- Who will you go with?

3) The student answering uses the flashcard the questioning student's holding as the answer, and visa versa,

4) Once completed, the pairs split and find a new pair asking the same questions, but will be faced with different flashcards to answer.

(The pair asking the question will correct grammar if necessary because They've perfected grammar for own picture vocab).



Where are you going?



I'm going to...
the beach!

Presentation



Activation: In pairs, make your own dialogue for presentation to class.

Task 5.2 – Speaking Group Critique/Notes

- ▶ The group have done a very thorough job in terms of presenting a set of slides that follow the ABLE model and provide comprehensive coverage of the various assessment metrics they will use (though with some almost direct copy/pasting).
- ▶ The activities show a good progression from closed to more open tasks with greater productive demands, and also demonstrate an interesting range and variety of tasks to stimulate learning via different methods.
- ▶ The overall content is very good and quite comprehensive, though if I am being critical I think the somewhat haphazard presentation style of the information/slides in terms of their formatting could be improved.

Task 5.3 – Testing Reading (My group)

Slides by:

Natalie Balderstone

Rachel Moimane

Aren Tyr

(Michael Gibson was absent on
both Weds 9th & Thurs 10th)

Performance outcomes

Students does	Assessment
Reads the text, highlights new words and make appropriate notes.	Listen to the students while they read the text, make notes and record the audio.
Writes down the correct name of the animals that people can be frightened of in the provided space.	Demonstrate the activity and ask students question to check their understanding. Record the grades.
Closed focused questions-reads the text and completes the sentences.	Check how many questions are answered correctly, mark and record the results.
Answers the True/False Questions, using information from the text.	Grade students on answers provided and record.
Reads the questions, then practices asking and answering them with a partner.	Record for the language accuracy, reading and interviewing skills.
Student describe an answer to the teacher and give reasons.	Grade the correct answers and allocate scores.
In groups of four, two sets of pairs must role-play the scenario depicting what is on the card to the other pair.	Grades the role-plays for accuracy and overall presentation.

Activity 1

(warmer)

Write the correct name of the animals that people can be frightened of.

mouse

spider

bat

pig

insect

snake













Activity 2

Read the text and complete the sentences.

1. **Musophobia** is the fear of _____ .
2. **Entomophobia** is the fear of _____ .
3. **Swinophobia** is the fear of _____ .
4. **Arachnophobia** is the fear of _____ .
5. **Ophidophobia** is the fear of _____ .
6. **Cynophobia** is the fear of _____ .
7. **Alektorophobia** is the fear of _____ .

Activity 3

Read the text and circle either True or False.

- | | | |
|--|-------------|--------------|
| 1. When you are very scared of something it's called a phobia. | <i>True</i> | <i>False</i> |
| 2. Scarlett Johansson is scared of snakes. | <i>True</i> | <i>False</i> |
| 3. Swinophobia is the fear of pigs. | <i>True</i> | <i>False</i> |
| 4. Some people are scared of cats. | <i>True</i> | <i>False</i> |
| 5. Orlando Bloom has a phobia of dogs. | <i>True</i> | <i>False</i> |
| 6. Alektorophobia is the fear of owls . | <i>True</i> | <i>False</i> |
| 7. There is a town in England where spiders fall from the sky. | <i>True</i> | <i>False</i> |
| 8. The spiders and their webs are blown by the wind. | <i>True</i> | <i>False</i> |

Activity 4

Pair work.

Take turns asking and answering these questions with a partner:

1. Are you scared of spiders?
2. Are you scared you scared of snakes?
3. Which do you think is scarier, pigs or mice?
4. Is there anything else you are scared of?
5. Do you think it's silly to be scared of animals like cats or dogs?

Activity 5

There is a town in Brazil where spiders fall from the sky. There are many spiders up in the trees, on buildings, and on telephone wires. They build giant webs. When it is very windy the webs and all of the spiders can get blown away. This is when they seem to rain from the sky.

So many people in this town in Brazil have a phobia of spiders. They think the spiders are very scary. They are frightened by them.

Do you remember where the text talked about spiders that ‘fall from the sky’ in Brazil?

Watch the YouTube video online here <https://www.youtube.com/watch?v=6qpsQg9pWJo>.

1. Write down **3** words that describe how you feel when you watch the video. What do you feel? So think of **three words** and write these down.
2. Look at the last part of the text again. Why do you think spiders are falling from the sky? Write down your answers.
3. Talk to the person next to you and compare your answers for Q1 and Q2 above.
 - i) What words did they write down? Did they think of different words?
 - ii) Why do they think that spiders are falling from the sky?

Activity 6

1. Students gather in groups of four; two sets of pairs.
2. In their pairs, they must role-play the scenario depicted on the card to the other pair.

A

You and your friend are eating dinner in the kitchen after a busy day out shopping.

Whilst you are eating a dinner, your friend sees that a giant spider is walking across the kitchen floor!

What will you do?

B

It is a hot summer's day and you and your friend are out camping in the woods.

You are both relaxed after just having eaten a nice meal that you cooked on a fire, and are starting to fall asleep in your tent. As you are both falling asleep, a large pig appears, and runs straight into your tent!

What will you do?

C

You and your friend were sitting by the river when you noticed a big brown cat go running by at great speed.

You chased after the cat and found it sat down near a big tree.

Suddenly a mouse appears from a hole by the tree and starts running straight towards you!

What will you do?

D

You are both out in the jungle. You are travelling on a big adventure to an ancient city. There are many trees and the jungle is quite dark. As you are walking, you see a big snake drop down from the tree onto the path in front of you!

What will you do?

Task 5.3 – Reading Group Critique/Notes (My group)

- ▶ Our slides are lacking content when compared with the speaking group, for example, particularly in terms of the *assessment metrics*.
- ▶ I struggled to try to get everyone to work together more coherently, as I felt there was a general tendency for each person to go in their own direction and insufficient communication as to the performance objectives we were seeking to assess. This is, I suppose, a fairly normal state of affairs in a new group dynamic as establishing a good workflow and communication pattern is not straightforward and takes time. In particular, on Thursday morning I tried to suggest that we were putting **too much focus on designing activities somewhat arbitrarily (i.e. material selection/construction)**, instead of **clarifying performance outcomes** which would then serve as the basis for selecting/designing activities. This to me is reflective of the fact that this approach is perhaps counterintuitive to teachers experienced with the traditional model, orientated around their personal approach and the planning of activities on this basis, with the outcomes at the end rather than the beginning.
- ▶ The response was that I was being over-analytical – which may perhaps have been true. I still feel we could have followed the ABLE model in terms of workflow more closely though.

- ▶ As a result of all these difficulties, I feel that the spread of activities and skill level required is too broad. The early activities are extremely rudimentary, before progressing up to considerably more difficult reading skills later on, a spread I feel is too large within the context of one lesson/one class; earlier activities are either likely to be too easy, or later activities will prove too difficult.
- ▶ The presentation of our slides is rather spartan though they are clear.
- ▶ I wrote activities 5 and 6 above and the issue we have is that the potential skill requirement/demand ramps up somewhat steeply with regard to our earlier activities. I feel this is a result of our lack of sufficient attention to an *outcome focused planning approach* rather than an activities designing approach, which after all was the intention of the exercise. I had to defer to the group majority in terms of our approach.
- ▶ The performance outcomes could be more expansive and more specific, and could also be split into formative and summative outcomes. When I suggested that they we should work to improve them, my group members responded that they thought they are good enough. I have presented my interpretation of summative/formative performance outcomes and my general approach with more detail in my grammar testing section below. This is the model I would have preferred us to follow.
- ▶ Some slides relating the performance outcomes and activities in the form of lesson stages would be beneficial to our work. I have put a set of these in to my grammar testing slides below. These could certainly be expanded further but I have had to compromise based on time constraints, though I think they are sufficiently comprehensive for their intended purpose.
- ▶ One of our assigned group members was absent throughout both days of jigsaw work, so **consequently there were only actually three of us to complete all the work on this task** (I note that some other groups had six members...). We could certainly have provided a more comprehensive set of slides with more members contributing, though the speaking group did an excellent job with just four group members.

Task 5.4 – Testing Writing

Slides by:

Jack Alexander Hollywood

Jason Kendall

Ursula Bradshaw

Damian Boshoff

Steven Cruz

Corey Prudhomme

Writing Group

Second Language Testing

Jack Alexander Hollywood, Jason Kendall, Ursula
Bradshaw, Damian Boshoff, Steven Cruz, Corey
Prudhomme

Table of Contents

1. Summative Outcomes
2. Formative Outcomes
3. Performance Outcomes
4. Warmer
5. Context setting & modeling
6. Assessment - Activity 1
7. Assessment - Activity 2
8. Assessment - Activity 3
9. Assessment - Activity 4
10. Assessment - Activity 5
11. Assessment - Activity 6
12. Assessment - Activity 7

CEFR Level - Writing and Grammar

CEFR Level	Language level	Testing activity
B1	Multiple sentence	<ul style="list-style-type: none">• Dialogues• Short Stories
B2	Paragraph	<ul style="list-style-type: none">• Letters• Essays

Scoring Rubric for Assessment Activity 7

The score for writing task (activity 7) will be out of 20, with points out of 5 being given for each of the following criteria:

- Grammar
- Lexical resource use and cohesion
- Following the outlined structure
- Coherence and Cohesion

0	1	2	3	4	5
Student has insufficient participation to evaluate OR No attendance OR Plagiarized	Very low participation. The student has not shown progress yet.	Below expectations for this level but the student has shown some progress.	The student meets the expectations for this level. Some frequent errors occurred. Some evidence of progress observed.	The student exceeded most expectations for this level. The student definitely showed progress and improvement.	Exceeded all expectations for this level. Very little or no mistakes. Ready for the next level.

Summative Outcomes

Students will be able to:

- Describe in writing a favourite photo and the surrounding events in the correct tenses, using the activities and examples as guidelines.
- Show understanding of target language input, both grammatical and lexical items by completing the given tasks in writing.

Formative Outcomes

- In small groups, **write** answers to questions about the pictures
- In pairs, **write** sentences in pairs about pictures A, B and C in the correct word order
- Individually **read** and **match** the text to the pictures
- Individually **fill in** the blanks with descriptive words
- In trios, **complete** the dialogue by **writing** hypothetical answers
- In pairs, **write** an imaginative ending to the story
- Individually **write** sentences comparing and contrasting the emotions of the people in the pictures
- Individually **circle** the correct preposition
- In pairs, **fill in** the blanks using the correct verb forms and **recognize** the writing structure **by matching**
- Individually, **write** a description about your own favourite photo

Performance Outcomes - Warmer

Student Does	Teacher Does	Assessment
Watches teacher model example of picture 1, listens and responds to teacher Discusses in small groups	Models example TS elicitation	T notes the general level of Ss
Looks at picture 2 and writes the answers down in small groups	Monitors the activity	T notes the general level of Ss' writing

Performance Outcomes - Context Setting and Modeling: Unscramble the sentences

Student Does	Teacher does	Assessment
Writes down the words in correct order in pairs	Models how to unscramble the words and monitors Ss' work	T checks the word order and takes notes
Individually reads and matches text to pictures	Models, monitors and explains any unknown lexical items	T facilitates the peer assessment of correct matching by swapping papers

Performance Outcomes - Activity 1: Adding words

Student does	Teacher does	Assessment
Individually fills in the blanks with descriptive words from the box beneath	Models and monitors	T elicits answers from the class and notes common errors.

Performance Outcomes - Activity 2: Complete the dialogue

Student does	Teacher does	Assessment
In trios look at the photo and discuss possible answers to complete the dialogue.	Models and monitors	T listens to S say their answers and notes any common errors.

Performance Outcomes - Activity 3: Finish the story

Student does	Teacher does	Assessment
Discuss in pairs possible story endings	Models and monitors	T listens and notes activity success
Write the ending to the story together	Monitors and provides feedback	T collects and scores papers out of 10 with minimum requirement of 5

Performance Outcomes - Activity 4: Compare and contrast

Student does	Teacher does	Assessment
Note the example and ask questions if any	Models and answers any questions	T checks level of understanding
Individually write their answers on the paper	Models and monitors	T listens to Ss read their answers aloud and notes any common errors

Performance Outcomes - Activity 5: Multiple choice

Student does	Teacher does	Assessment
Individually circles the correct prepositions	Models and Monitors	T facilitates the peer assessment of correct circling by swapping papers

Performance Outcomes - Activity 6: Gapfill and structure recognition

Student does	Teacher does	Assessment
Writes the correct verb in the blank in pairs	Monitors and Models	T makes note or video of the S completing the task
In pairs matches the correct sentence to the paragraph by writing the correct number in the box	Monitors and Models	T listens to S say their answers and notes any common errors

Performance Outcomes - Activity 7: Write your own photo description

Student does	Teacher does	Assessment
Reads the example out loud and answers the questions	Nominates Ss to read different sentences and corrects pronunciation	T records the reading audio and takes notes
Individually writes description of their own photo.	Monitors and assists when necessary	T collects and takes scores according to criteria rubric

Warmer



Look at these two photos. Then write about:

1. What was the photographer doing when they took the photo?

1. What were the people doing in the photo?

Context Setting and Modeling

Picture A



Picture B



Picture C



Context Setting and Modeling - Unscramble The Sentences

making He was speech a .

1. _____

They song composing a were .

2. _____

musical playing instruments They were .

3. _____

with wall a hitting was She the rock .

4. _____

as the United president He States of resigning was the of .

5. _____

the She shouting was crowd with .

6. _____

Core Text

Match the pictures to the photos.

*Harry Benson spent 50 years taking dramatic and memorable photos for national newspapers and magazines. Here he talks about three of his best-known pictures.*¹

1

I took this picture on August 9th 1974. He was saying goodbye to his cabinet and the White House staff after the “Watergate” scandal. All his family, were standing round him. You can see from their faces what they were feeling.

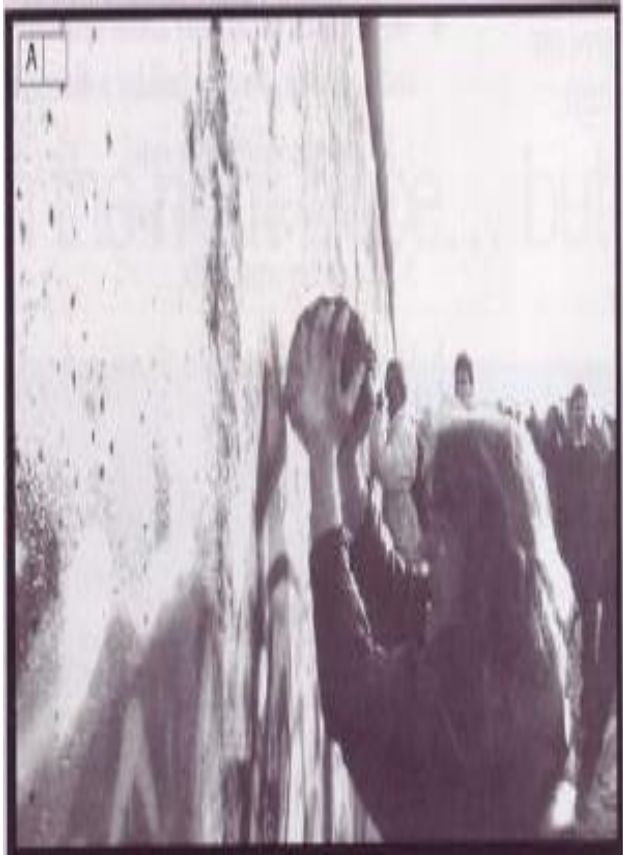
2

In 1989, I was working in London on a story when suddenly I heard the news that Russians were planning to make Berlin an open city. So I got on a plane. When I arrived in Berlin many people, young and old, were attacking the wall with stones. The woman in the photo was shouting, ‘I hate it, I hate it!’.

3

When I took this photo we were in a hotel room in Paris in 1964. John and Paul were at the piano and at first nothing much was happening, but suddenly they became completely focused. First the melody came, and then the words. *‘Baby’s good to me you know, she’s happy as can be you know...’*. They were composing their song, ‘I feel fine.’

Activity 1 - Adding Words

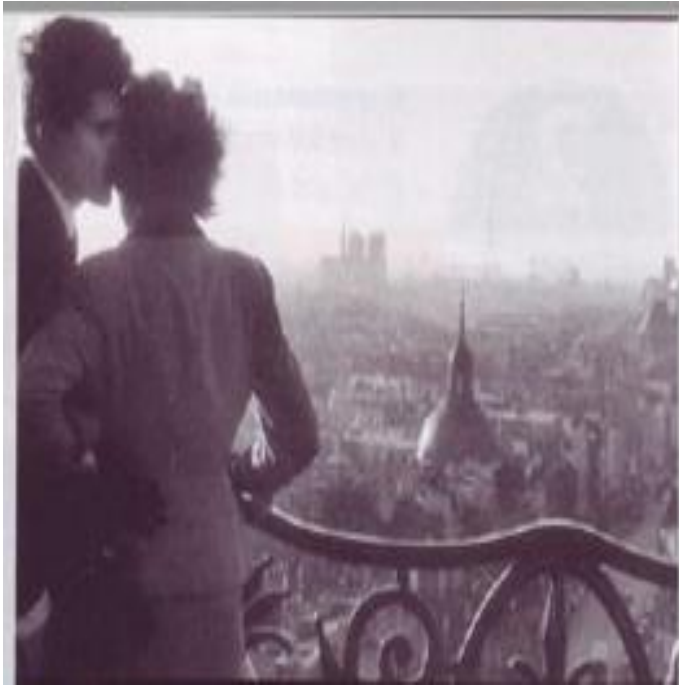


Add descriptive words to the text from the box below:

In 1989, I was working in _____
London on an _____ story
when suddenly I heard the news that
Russians were planning to make Berlin
an open city. So I got on a
_____ plane. When I arrived
in Berlin many people, young and old,
were _____ attacking the wall
with stones. The _____
woman in the photo was shouting, "I
hate it, I hate it."

angry | private | violently | exciting | beautiful

Activity 2 - Complete the dialogue



What do you think they were saying to each other?

Jacques: "What do you think of the view?"

Marie: "_____"

Marie: "Where do you think that car is going?"

Jacques: "_____"

Jacques: "What do you think that shop is selling?"

Marie: "_____"

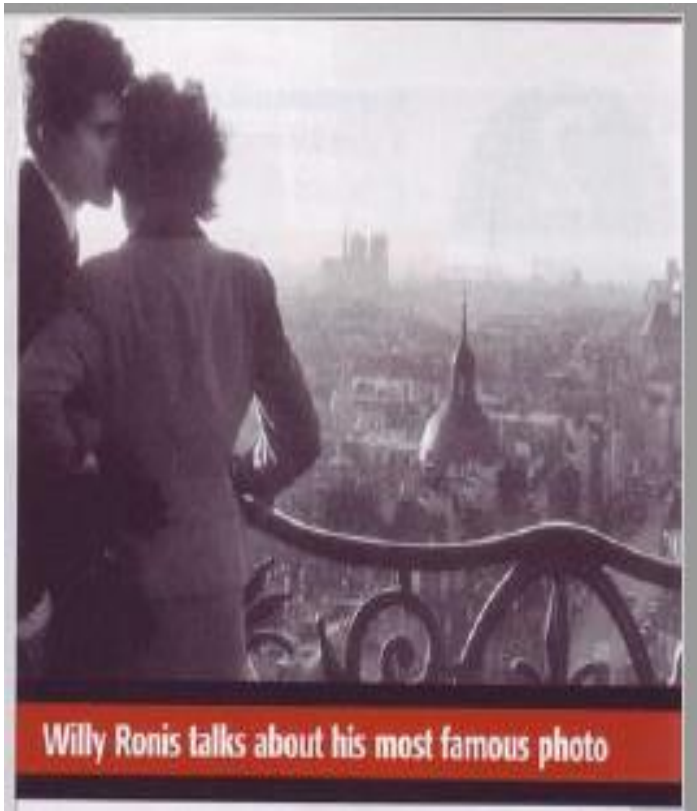
Marie: "What should we have for dinner tonight?"

Jacques: "_____"

Jacques: "Will you marry me Marie Le Blanc"

Marie: "_____"

Activity 3 - Finish the story



It was March 1957 and I was taking photographs in Paris. One afternoon I went up the Bastille and I saw two lovers on a balcony. They were standing very near each other. They were talking. I took just one photo and they didn't hear me. I called it *Lovers at the Bastille*. Luckily for me this photo became very popular in France. Soon posters and postcards with my picture of the two lovers were everywhere. But I never knew who the two young people were. They never contacted me.

30 years later I had an exhibition of my photos in Paris. I was talking to some friends when suddenly a man came up to me and said, "I know your two lovers. They live near here. I can take you there if you want." I immediately decided to go and meet them. This was their story.

Complete their story - What happened to Jacques and Marie after the picture was taken?

Activity 4 - Compare and Contrast

Compare the emotions of the people in the 4 photos using the words given.

Here is an example:

*Example: woman/Berlin/couple/Paris (**angry/loving**)*

The woman in Berlin was angry, while the couple in Paris were loving.

1) Men/hotel/family/America (**creative/disappointed**)

1) Couple/Paris/men/hotel (**romantic/focused**)

1) President/Washington/woman/Berlin (**ashamed/furious**)

Activity 5 - Grammar activity - Multiple Choice

Choose the correct preposition of time to complete the sentence

1. President Nixon left the White House ___ August 9th 1974.

- a) in b) at c) on d) for

2. The Beatles were ___ a hotel room ___ Paris ___ 1964.

- a) on b) at c) in d) over

3. Willy Rollins took the photo (1) ___ the balcony (2) ___ March 1957.

- (1) a) at b) in c) on d) under

- (2) a) in b) on c) at d) above

4. The young couple went up to the balcony ___ 3.00 in the afternoon.

- a) in b) on c) at d) to

5. Henri and Marinette see each other everyday, ___ home and ___ work.

- a) at b) in c) on d) with



Activity 6 - Gap Fill

Fill in the blanks with the missing verbs in their correct form:

captu re	sail	imagi ne	cook	catch	taste	look
-------------	------	-------------	------	-------	-------	------

1. One of my favourite photos is of a man _____ fish.
2. I _____ the photo in the summer of 1994 when I was on holiday in Africa.
3. We were in Zanzibar, and we were _____ round the islands. We _____ some fish and then landed on one of the islands, and our guide made a fire and cooked the fish.
4. Richie, the guide, was a wonderful, friendly person - he became a good friend. He was also a fantastic cook. I can't remember exactly what he was cooking in the photo, but it _____ delicious!
5. I keep this photo on the wall in my studio. On a grey, rainy day in London, I just have to _____ at it and then memories of a wonderful adventure come back, and I can _____ that I'm in sunny Africa again.

Activity 6 cont. - Writing Structure Recognition

Match the content sentences with the paragraphs 1 - 5:

What was happening when you took the photo?

☐

Where do you keep it? Why do you like it?

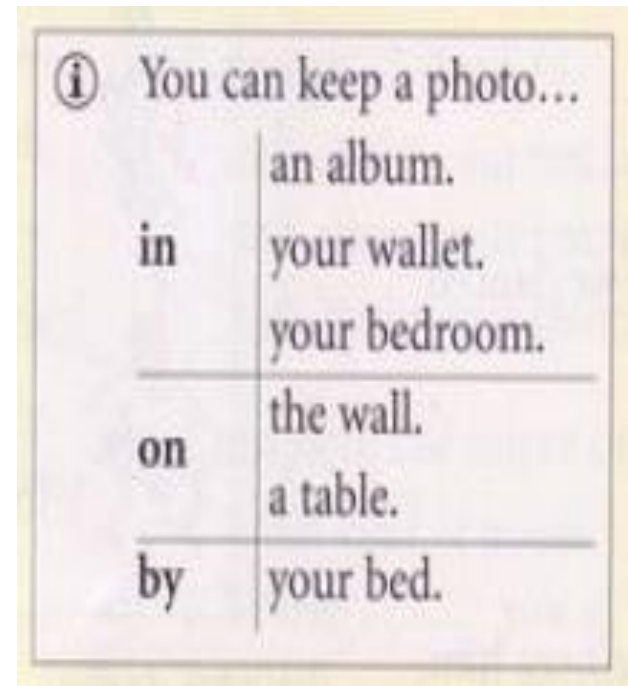
☐

What's your favourite photo?

☐

Who took it? When? Where?

Tell me more about who or what is in the photo.

☐☐

Activity 7 - Write your own photo description



Write about your own favourite photo using the correct tenses. Here is an example:

One of my favourite photos is of a squirrel “photobombing” a holiday photo.

I took the photo in 2012 when I was in Canada with my friends Jack and Ursula.

We were on holiday in the Rocky Mountains, and we were hiking around a lake. We found a beautiful view of the lake and decided to stop for a rest.

I set up the camera and was getting ready to take the photo. As I was taking the photo, a squirrel came into the view and looked at the camera. I took the photo at just the right time and it looks so funny!

I keep this photo on the wall of my office at work. So if I am having a tough day, I can just look at this funny picture to put a smile back on my face.

Task 5.4 – Writing Group Critique/Notes

- ▶ The group has put together an interesting and challenging selection of activities for assessing writing, with a good progression.
- ▶ The group has also clearly identified performance outcomes at each stage, provided a set of summative outcomes, and have also showed the assessment procedure for all of these outcomes. The summative outcomes are not as clearly presented/strongly written as the formative outcomes.
- ▶ Apart from some minor typographical/grammar errors, the slides in general are clear and well presented (some did not copy/paste very well into my slides here, because I think the PPT file originated from a Google Docs template with different page/slide size to the ones I am using here).

Task 5.5 – Testing Grammar

(My work alone)

- 1. Class profile**
- 2. Summative outcomes**
- 3. Formative (Performance) outcomes**
- 4. Lesson stages**
- 5. Sample script**
- 6. Activities**

1. Class Profile

CEFR Level: A1 (Young learners)

- *Basic repertoire of words and simple phrases*
- *Limited control of a few simple grammatical structures*
- *Can manage very short, mainly pre-packaged utterances*
- *Can link words with very basic linear connectors like 'and' or 'then'*

2. Summative Outcomes

Students will be able to:

- *Identify four nouns that relate the word with their specific object (**house, car, tree, and cat**)*
- *Spell and pronounce these four nouns*
- *Use a basic adjective (**big, fat, green**) to describe features of its corresponding object*
- *Spell and pronounce these three adjectives*
- *Demonstrate some partial understanding of the structure of a very basic sentence*

3. Formative (Performance) Outcomes 1/3

Students does:	Assessment:
<ul style="list-style-type: none">• <i>Matches</i> the word (house, car, tree, and cat) with the corresponding picture with 85% accuracy.	<ul style="list-style-type: none">• <i>Demonstrate the activity and ask students questions to check their understanding. Monitor and mark the worksheet.</i>
<ul style="list-style-type: none">• <i>Locates</i> the words in the word search by circling the word and crossing off from the word list with 95% accuracy.	<ul style="list-style-type: none">• <i>Demonstrate the activity if needed and ask students questions to check their understanding. Monitor and mark the worksheet.</i>
<ul style="list-style-type: none">• <i>Unscrambles</i> the word (house, car, tree, and cat) by writing it in the corresponding space with 85% accuracy.	<ul style="list-style-type: none">• <i>Demonstrate the activity and ask students questions to check their understanding. Monitor and mark the worksheet.</i>

3. Formative (Performance) Outcomes 2/3

Students does:

- *Reads the four questions and **circles the picture** in each case that has the corresponding adjective as applied to the object with 80% accuracy.*
- *In pairs, practice reading each **sentence aloud** with 50% accuracy (sound/pronunciation).*
- *Completes the gap fill by **filling in** the four blank spaces in the sentences with the appropriate word with 85% accuracy.*
- *In groups of three, **moves** the word cards around on the table to make correct sentences with 65% accuracy.*

Assessment:

- *Demonstrate the activity and ask students questions to check their understanding. Monitor and mark the worksheet.*
- *Monitor and listen to student-student interactions. Students will provide a form of peer review by listening/scaffolding.*
- *Demonstrate the activity and ask students questions to check their understanding. Monitor and mark the worksheet.*
- *Demonstrate the activity and ask students questions to check their understanding. Monitor and listen to student-student interactions.*

3. Formative (Performance) Outcomes 3/3

Students does:

- *In groups of four, each student gets secretly shown a flashcard which they will then mime. The other three students will then **name** the object their fellow student is **miming** by **guessing**, and **speak aloud** any other words or sentences that they can produce that might plausibly **describe** it with 45% accuracy.*

Assessment:

- *Monitor closely and listen to student-student interactions and speech. Record audio clips of group work for later assessment/review.*

4. Lesson stage: Warmer

Student does	Teacher does	Assessment
<p>Listens to teacher.</p> <p>Responds to greeting by repeating “hello” back to the teacher.</p>	<p>Teacher models/provides greeting.</p>	<p>Teacher notes general level of competence with greets and names.</p> <p>Teacher notes general level of attention.</p>
<p>Listens to teacher.</p>	<p>Teacher introduces the topic by presenting the script & pictures on whiteboard/display screen.</p> <p>Teacher reads the script aloud three times:</p> <p>1.) Very slowly with lots of miming/gestures and by pointing at each picture in turn</p> <p>2.) Slowly with miming once more</p> <p>3.) Slowly with no miming</p>	

4. Lesson stage: Context & modelling

Student does	Teacher does	Assessment
Listens to teacher.	Teacher presents each flashcard of the noun in turn (house, cat, car, tree) and speaks clearly.	Teacher notes general level of concentration whilst introducing each word.
Listens and repeats aloud.	Teacher does a whole class choral drill with the four flashcards.	Teacher listens carefully for pronunciation errors and remodels/repeats drill as necessary until correct.
Listens to teacher.	Teacher presents each flashcard of the adjective in turn (big, green, fat) and speaks clearly.	Teacher notes general level of concentration whilst introducing each word.
Listens and repeats aloud.	Teacher does a whole class choral drill with the three further flashcards.	Teacher listens carefully for pronunciation errors and remodels/repeats drill as necessary until correct.
Listens to teacher.	Teacher introduces the adverb (very) by showing matched pairs of flashcards (a big & very big house, a fat & very fat cat, a red and very red car).	Teacher notes general level of concentration whilst introducing each word.
Responds to teacher questions by pointing/speaking toward the correct card.	Teacher uses Q&A to whole class/individual students to get them to identify which card out of the pairs is the “very” version. Teacher tests understanding by using two new cards to check whether they can generalise the concept.	Teacher listens to students suggested answer and notes struggling students.

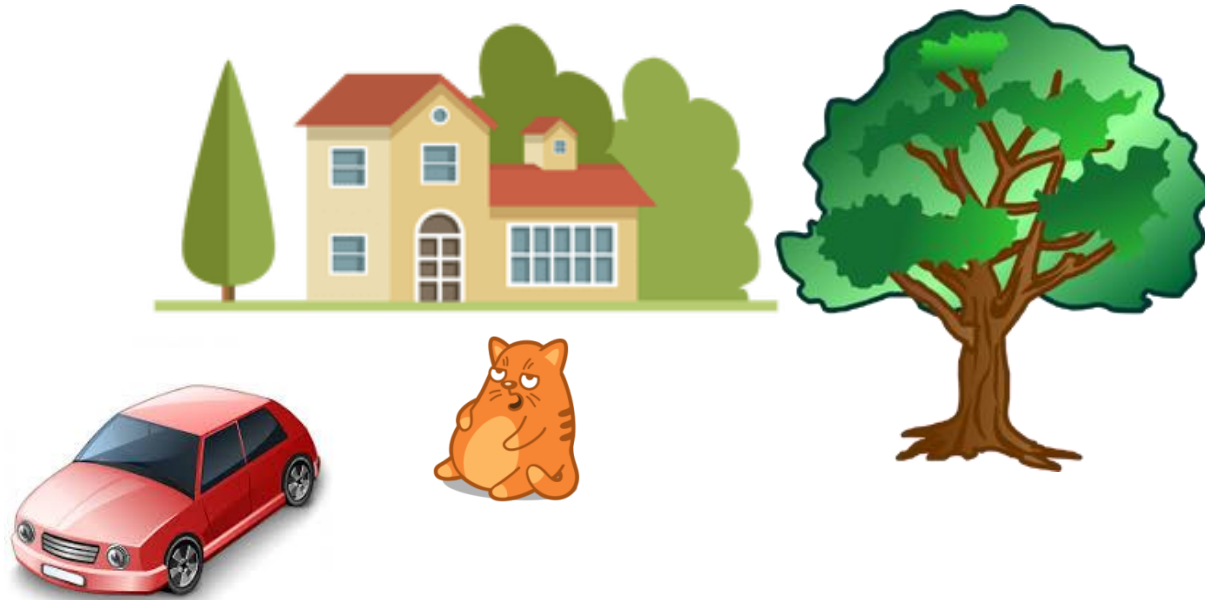
4. Lesson stage: Body

Student does	Teacher does	Assessment
Matches the words with pictures (activity 1).	Teacher circulates and monitors, offering student led discovery where needed.	Teacher notes any students struggling to identify the words.
<p>Circles the words in the word-search (activity 2).</p> <p>Student speaks aloud to teacher.</p>	Teacher circulates and monitors, offering student led discovery where needed. Teacher also uses opportunity to check students individual pronunciation of words by asking them to speak particular words aloud.	Teacher notes any students struggling to identify the words and listens for pronunciation on elicited words.
Unscrambles the words and fills in the sheet (activity 3).	Teacher circulates and monitors, offering student led discovery where needed.	Teacher checks for correct spelling.
<p>In pairs, identifies the correct picture by circling it (activity 4). Each student then speaks the sentence describing the object aloud to their partner.</p> <p>Student answers questions from teacher.</p>	Teacher circulates and monitors, offering student led discovery where needed. Teacher asks additional questions based on flashcards they are holding to see if they can abstract the concept.	Teacher checks correct identification and listens carefully for pronunciation errors. Teacher asks further questions based on flashcards to check degree of comprehension.
<p>Students fill in the gap fill sheet with the correct term (activity 5).</p> <p>Students read sentences (or words, if struggling) aloud to the teacher.</p>	Teacher circulates and monitors, offering student led discovery where needed. Teacher asks students to read individual completed sentences (or just words) aloud to help them relate the words with their sounds (phonology).	Teacher checks that they have identified the correct grammar item. Teacher notes level of overall prosody of each student.
Students rearrange the word cards in groups of three into the correct order (activity 6).	Teacher circulates and monitors, offering student led discovery where needed.	Teacher notes contributions/input of individual students into the group activity. Teacher listens to individual student-student speaking communication.

4. Lesson stage: Plenary/Communicative Activity

Student does	Teacher does	Assessment
<p>Acts/mimes out their flashcard to the other student who must guess the object (activity 7).</p> <p>Students speak aloud and guess to what object the student is miming. Students also offer input of further words/sentences that might serve as descriptive terms with elicitation/assistance from the teacher.</p>	<p>Teacher circulates and monitors through each group in turn, offering student led discovery where needed. Teacher also uses opportunity to elicit further vocabulary terms and potentially elicit complete basic sentences from more advanced students.</p>	<p>Teacher notes contributions and engagement to group activity. Teacher notes level of speaking competence and ability of individual students.</p> <p>Students provide feedback by altering miming/responding to other students guesses/input.</p>

5. Sample script



There was a **big house**. Around the **house** there was a **red car**, a **fat cat**, and a **very big green tree**.

6. Activities

Activity 1

Picture match

Match the word with the picture.

Car

Cat

Tree

House



Activity 2

Word search

Can you find all the words?

E	E	R	T	S	D	D	V	K	T
K	M	N	C	Y	H	G	H	M	O
U	F	P	C	A	R	O	Y	R	R
L	V	E	D	G	U	E	C	B	E
G	R	L	I	S	Z	A	V	Y	D
Y	R	B	E	T	T	C	M	H	X
Q	K	E	N	U	D	X	M	H	V
P	W	T	E	W	A	J	M	B	V
O	A	I	T	N	Y	Z	B	Q	Z
F	C	O	T	Q	R	O	H	P	U

HOUSE

TREE

CAT

CAR

RED

BIG

VERY

FAT

GREEN

Activity 3

Spelling Scramble

What are these words? Unscramble, write, and match.

Scrambled Word	Word:
cta	
rtee	
rac	
esuho	



Activity 4

Adjectives 1/4

Which cat is fat?



*Practice
reading:*

The cat is fat.

Activity 4

Adjectives 2/4

Which house is big?



*Practice
reading:*

The house is big.

Activity 4

Adjectives 3/4

Which car is red?



*Practice
reading:*

The car is red.

Activity 4

Adjectives 4/4

Which tree is very big?

Which tree is big?



*Practice
reading:*

The tree is very big.

Activity 5

Gap fill – complete the sentences using the four words at the bottom

*The cat was _____. The
_____ was big. There was a
very big green _____
outside. The _____ was
red.*

fat *house* *car* *tree*

Activity 6

Sentence scramble – Can you make these sentences correct by moving the words?

the cat fat was

house the was big

there very tree was a

big green outside

was car the red

Activity 7

Act/mime (1/4):



Activity 7

Act/mime (2/4):



Activity 7

Act/mime (3/4):



Activity 7

Act/mime (4/4):



Task 6 – How to validate tests

- ▶ Questions based on the slide set *Item Analysis and Validation* by Mark Leonard Tan, Verena Gonzales, Ann Creia Tupasi, and Ramil Cabañesas

a) Getting started

► ***1. What is the first step in preparing to analyse and validate test items?***

The first step is the production of a draft version of the test. This constitutes a 'try-out' phase. In essence a rough and ready prototype that is used as a 'field test' or experiment. In scientific parlance, rather like a laboratory sample.

► ***2. What are the first steps in preparing a test?***

So there are three essential steps; 1) The 'try-out' prototype phase, described above; next 2) *item analysis*, whereby each item that is on the test is individually statistically analysed using various metrics based on the results from the initial try-out field test; 3) *item revision*, whereby in light of the results of the statistical analysis any test items revealed to deficient are suitably modified.

The construction of a test therefore may be iterative, going through several (hopefully improved) versions, rather like the development of software, which slowly progresses from version '0.1' through to '1.0' which would constitute the first 'stable' or 'finished' version of the software.

▶ ***3. What are the two important characteristics to be studied?***

- ▶ The two characteristics that are studied on a per-item basis are i) **‘Item difficulty’** and ii) **‘Discrimination index’**.

▶ ***4. Describe what is meant by these characteristics.***

- ▶ **i) Item difficulty**
- ▶ Item difficulty is a ratio (typically presented as a percentage) representing how many students successfully answered the given item correctly. It is obtained by simply dividing the number of correct answers by the total number of students. It thus offers a crude metric for determining how ‘difficult’ the item was.

▶ ii) Discrimination index

- ▶ The item difficulty ratio (expressed as percentage) in itself is a limited metric. It also does not indicate the actual level of difficulty of an item *in and of itself*. For example, a particular student may have been unable to provide a correct answer simply because they were completely unfamiliar with the material, even if it in actuality it was 'easy' item/question; the item per se was not too 'difficult'. So we therefore need a method of determining the true difficulty of a particular item.
- ▶ We can state that the effect of 'difficult' items is that they tend to divide students into a roughly binary division of those who *do* know and those who do *not* know. Conversely, 'easy' items tend to offer no separation at all.
- ▶ What we need is a metric that allows us to determine *to what degree* an item can differentiate between the two groups of students (those that 'know' vs. those who do not know) across the student cohort as a whole.
- ▶ The discrimination index serves this function.

► 5. *How is item difficulty calculated?*

- The difficulty ratio is obtained by dividing correct responses from the total responses of the cohort. A concrete example. Suppose we have a total of 117 students who responded to a test item, and 72 of these responses were marked as correct. Therefore:

$$\frac{72}{117} = 0.62$$

- So 62% of students delivered a correct response. In this case, this suggests that this item is of suitable difficulty, if we adopt this statistical schema based on acceptable ranges:

Difficulty Index	Interpretation	Action
< 0.25	Too difficult	Revise or discard
> 0.25 & < 0.75	Suitable difficulty	Retain
> 0.75	Too easy	Revise or discard

▶ ***6. What is the problem with this type of difficulty index?***

- ▶ As per my explanation of the discrimination index above, the item difficulty ratio (expressed as percentage) in itself is a limited metric. It does not guarantee a meaningful indication of the actual level of difficulty of an item *in and of itself*. It does not accurately differentiate between those that answer incorrectly due to insufficient or inadequate comprehension versus those who answer incorrectly simply due to complete unfamiliarity or ignorance of the material. In the latter case, even an easy item would lead to an incorrect (or entirely missing) answer.
- ▶ In other words, it offers no information regarding what I would like to term in my own words as the ‘granulation of comprehensibility’. At the low end of this scale is incomprehension due to unfamiliarity; at the high end is incomprehension due to factors relating to complexity, cognitive demands and prerequisite language knowledge required.
- ▶ The Fulcher and Davidson text describes this relationship in terms of that fact that the calculation of the value for the nominal ‘item difficulty’ depends as much on the ability of the test takers as it does any intrinsic difficulty of the item; this ‘sample dependence’ means that with a different set of test takers/students, the apparent item ‘difficulty’ value would change. The difficulty index therefore does not give us a reliable independent metric for assessing individual items; an ideal item ‘difficulty’ calculation method would exhibit sample *independence*, such that the nominal difficulty rating would be closely consistent regardless of sample used.

- ▶ We can glean a much more meaningful insight into the true measure of an item's difficulty by considering the *correlation* between correct answers to the item when indexed against the overall test results/performance across the student cohort. This will tell us with much greater precision as to its *actual* difficulty since outlier cases due to *unfamiliarity with this specific item* will become statistically insignificant and normalised given a sufficiently large student sample/cohort (student missed that day in class; student didn't revise that particular example; student ignored that specific topic due to boredom; etc.).
- ▶ Put another way, if our objective is to discriminate *between higher and lower ability test takers*, we need a method to quantify this, and the raw 'item difficulty' ratio does not adequately meet this requirement since it does not provide sufficient information.

b) Difficulty & discrimination

- ▶ ***7. How do we develop an index of discrimination? (Illustrate with examples and formula).***

- ▶ Given the limitations with item difficulty described above, we therefore need to calculate an 'index of discrimination' in order to give us an insight into how difficult an item actually *is* for our test audience, and whether we can therefore use it as an appropriate item for discriminating between higher and lower ability students. In other words, to tell us how well our test students know the material being tested.

- ▶ One of the simplest methods of obtaining an index of discrimination is by subtracting the item difficulty value of the upper quartile (i.e. the top 25% of test takers) from the lower quartile (bottom 25%). Here is a concrete example:
 1. Suppose the item difficulty for the upper quartile (DU) is 0.82 – i.e 82% of those students in the top 25% of the test takers answered the item correctly.
 2. Next, suppose the item difficulty for the lower quartile (DL) is 0.16 – i.e. 16% of those students in the bottom 25% of the test takers answered the item correctly.
 3. We then calculate the index of discrimination by the following formula:

$$DU - DL = \text{Index of discrimination}$$

i.e.

$$0.82 - 0.16 = \mathbf{0.66}$$

So our index of discrimination in this case is 0.66.

▶ ***8. What is the range of the index of discrimination and what does it mean? (Illustrate with examples).***

- ▶ The range for this occupy can mathematically occupy -1.00 (DU = 0, DL = 1) to +1.00 (DU = 1, DL = 0).
- ▶ An index of discrimination that has a negative value raises a red flag for the item under consideration, since it means that the students who have scored best

- ▶ in the test overall also happen to exactly the group of the one's who have performed worst overall in this particular item. In short, we have a form of *negative discrimination*. Such a case is not just a mathematical possibility, but can occasionally in fact occur in reality, in the case of badly designed test items, particularly poorly designed multiple choice questions.
- ▶ How might this occur?
- ▶ One plausible situation is with an ambiguous multiple choice question where more than one response is correct when subjected to more detailed scrutiny. Lesser able students, whose comprehension hasn't advanced beyond mere memorisation and 'rote' learning may immediately choose the initially 'correct' answer on this basis, with no awareness of the underlying ambiguity. More advanced students, whose comprehension and power of analysis is greater, may spot the ambiguity, and now faced with a dilemma, have to make a judgement call as to which 'correct' answer to select. I would suggest that such ambiguity is an easier trap to fall into than may be readily apparent, particularly when dealing with multiple choice questions (on whichever subject) that deal with higher level skills or more advanced subject matter.
- ▶ Worst of all is the case where the 'correct' answer in fact is incorrect when considered at a higher level of understanding. In a case where discrimination values approaching -1.0 are occurring, that is strong evidence to suggest there is a very serious flaw with the test item in question.

- ▶ The converse, of course, is equally true. Values approaching +1.0 would suggest a near perfect level of discrimination; the most able students all get the correct answer, whilst the least able get it wrong. The item provides an excellent degree of discrimination and suggest that the test item is nearly perfectly designed in the context of the test.

▶ ***9. What is the rule of thumb for determining which items to i) discard ii) revise iii) include?***

- ▶ In practice, rather like the difficulty index, we have a statistical schema that serves as general guide as to acceptability:

Index Range	Interpretation	Action
< -0.5	Highly suspect discrimination	Discard
> -0.5 & < +0.45	Non discriminating	Revise
> +0.45	Good discrimination	Include

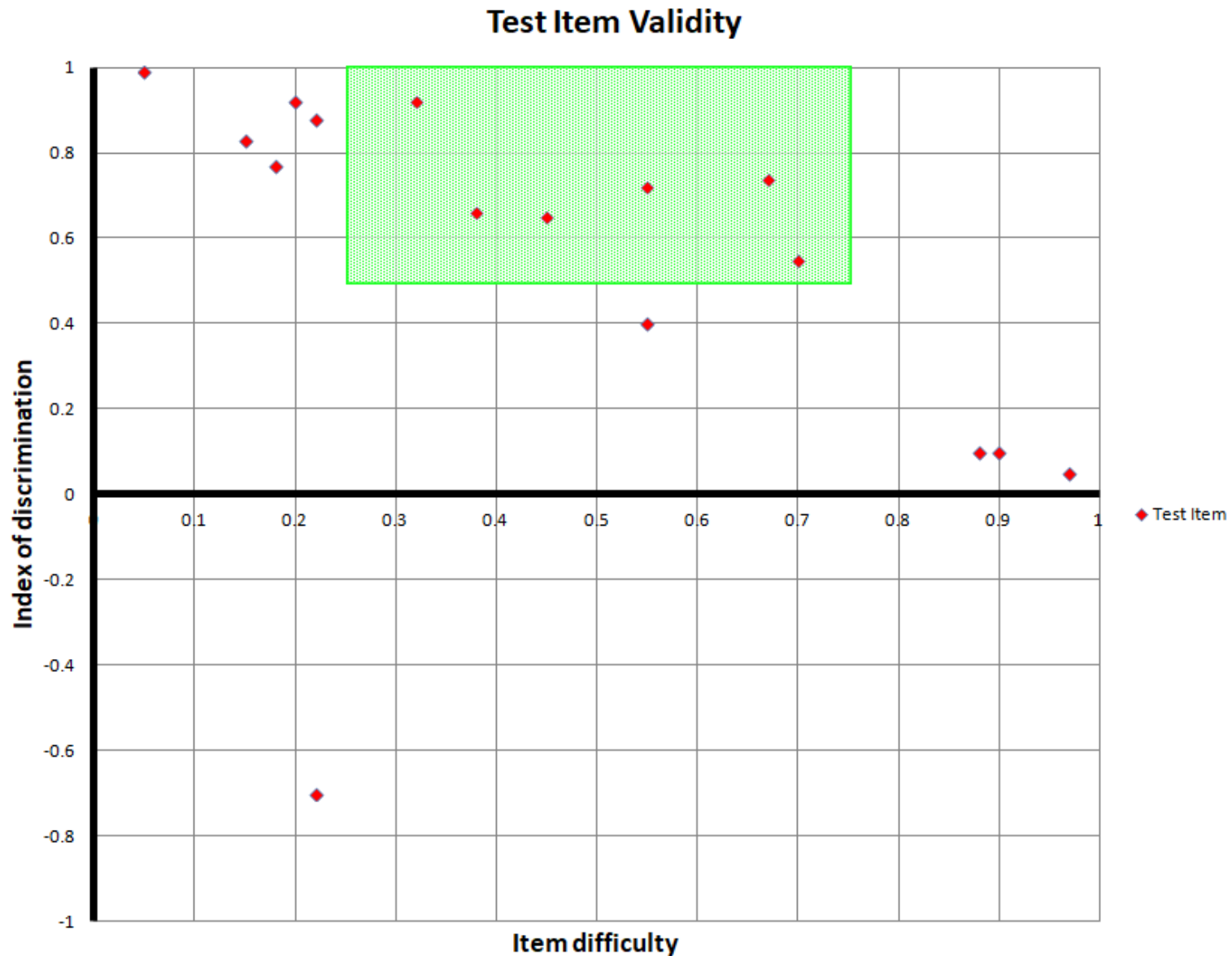
- ▶ The highly suspect (i.e. large negative values approaching -1.00) and extremely good discrimination (i.e. Large positive values approaching +1.00) have been discussed above. That leaves the 'non-discriminating' items whose index of discrimination falls between -0.5 and +0.5. These are items which fail to adequately discriminate, so on that basis need to be redesigned so they discriminate more acutely.

► ***10. Graph this for a multiple choice question showing difficulty index and index of discrimination***

- Consider the following set of test scores for our test items:

Test Item	Difficulty	Index of discrimination
A	0.15	0.83
B	0.67	0.74
C	0.88	0.1
D	0.22	-0.7
E	0.55	0.40
F	0.20	0.92
G	0.70	0.55
H	0.90	0.10
I	0.45	0.65
J	0.55	0.72
K	0.05	0.99
L	0.18	0.77
M	0.38	0.66
N	0.32	0.92
O	0.22	0.88
P	0.97	0.05

- If we now plot these, we get the following chart:



- The items *inside* the green box are the valid test items. They are both suitably difficult *and* they also successfully discriminate among students.

If we look at the data again, the **green** highlighted items are the **valid** test items. The **red** items are too *difficult* in general, and the **blue** items are too *easy*. Item E in **purple** is the correct overall difficulty, but it does not sufficiently *discriminate* based on student's knowledge of the material so needs improving. All of the items that are therefore *not* green would require either revising or simply discarding (i.e. **A, C, D, E, F, H, K, L, O, and P**). Item **D** is **highly suspect**; not only is it too difficult, it also *negatively* discriminates, which suggests that it is actually completely confusing our best testing students due to some intrinsic ambiguity.

Test Item	Difficulty	Index of discrimination
A	0.15	0.83
B	0.67	0.74
C	0.88	0.10
D	0.22	-0.70
E	0.55	0.40
F	0.20	0.92
G	0.70	0.55
H	0.90	0.10
I	0.45	0.65
J	0.55	0.72
K	0.05	0.99
L	0.18	0.77
M	0.38	0.66
N	0.32	0.92
O	0.22	0.88
P	0.97	0.05

c) Item difficulty statistics

▶ *11. University of Michigan item statistics?*

- ▶ The metrics described above give us various measures for assessing the validity/suitability of individual test items. *Index of difficulty* tells us in general the proportion of students in total that got it right/wrong; 0.85 means 85% of students answered correctly. *Index of discrimination* tells us the difference in proportion of correct answers between students in the top quartile versus those in the low quartile. For example, a high figure such as 0.80 would indicate that it discriminates well in that the best overall students in the top quartile managed to answer the item correctly, whilst the weakest overall students did not.

▶ *12. More sophisticated item analysis?*

- ▶ We are therefore interested in items that are the correct level of difficulty (not too hard, not too easy) *and* that also discriminate well, such as the items we identified in green in the above example on the previous slides. The combination of these two metrics together gives us a much better analysis/insight versus simply looking at how many students answered it correctly.

▶ *13. Benefits of item analysis?*

- ▶ There are three main benefits of item analysis.
- ▶ 1. It allows us to revise and improve the test. If we view the test as a whole as an evolving entity, it allows us to continually refine and improve the test for future cohorts of students. Questions will generally be “about right” in terms of difficulty and discrimination, meaning that the students that work hardest should in principle not only manage to answer most (though not all) questions, but that the questions will have an overall consistent trend where they are usually answered correctly by the best performing overall students, and incorrectly by the worst performing overall students (i.e. no suspect questions!).
- ▶ 2. It provides data upon which students can improve their learning.
- ▶ 3. It provides useful information for which the class can discuss the results of the test.

d) Validation

▶ *14. What is validation?*

- ▶ After having analysed the items in the test by combining the two main metrics described above (item difficulty and index of discrimination), we can make suitable modifications to the offending test items (revise or discard/create afresh), until all our *items* now meet our standard of validity. *Validating* refers to considering this entire test as whole in terms of collecting evidence that supports its usefulness and reliability. It also refers to the *appropriateness* of the test as a guide for the teacher's future decisions regarding students based on their test scores. If the test scores *consistently* do not seem to tally with the apparent actual ability of students, that would suggest that the test is not necessarily appropriate for testing what we intend it to test (i.e. it appears to be a test for 'testing's sake' alone rather than acting as a reflection of the best students).

▶ *15. What are the three main types of evidence to be collected? (Describe in detail)*

1) Content-related evidence of validity

- ▶ As the name suggests, this refers to the content of the test.
- ▶ Is it appropriate? (e.g. If the test is on the area of mechanical friction of plane surfaces, then presumably we do not want a test item on Nietzsche's existentialism, to use a crude example).
- ▶ How comprehensive? Does the test cover all the main aspects of the area we are interested in? (i.e. does the test adequately relate to the syllabus or section of syllabus it is intended to).
- ▶ Does it have a logical relationship with the intended 'variable'? i.e. Do the questions accurately pertain to their intended items under scrutiny?
- ▶ Does the sample of questions that appears in the test drawn from the original question pool manage to broadly assess all the aspects of the material we are interested in assessing in a consistent manner?

2) Criterion-related evidence of validity

- ▶ Criterion-related validity refers to using an external/additional form of evidence such that the scores can be cross-referenced or related in some way. The strength of the relationship between the original test and the external evidence, and the degree to which it can be used to predict future performance on the original test instrument undergoing validation indicates how effective this criterion will prove to be.

3) Construct-related evidence of validity

- ▶ How well does the test construct reflect the ability of individuals to perform a task directly related to the material the test is supposed to be assessing? How closely does the test reflect the psychological construct that is the object that the test is supposed to be scrutinising? Does the test provide a sufficient representation of the 'reality' of the material?

▶ ***16. What is the procedure for determining content validity?***

- ▶ The teacher(s) compile a Table Of Specifications (TOS) which provides a summary of topics/material the test will cover. From this TOS, they then write their test *objectives* (i.e. What is it that they want to identify that students will actually be able to *do* or demonstrate in the test).
- ▶ The TOS, objectives, and test items are then presented to at least two experts. The teacher(s) also present the experts with description of the test takers. The experts will then place a check mark next to the test items that they determine do **not** assess one or more of the test objectives. The experts will also place a check mark next to any **objective** that has **not** been tested by any item in the test.
- ▶ The teacher then revises and resubmits any flagged test items, and where necessary create new/additional test items to cover any objectives not currently being assessed.
- ▶ This back-and-forth process continues until all test items are approved and the experts agree that all of the objectives have been sufficiently covered by the test.

▶ ***17. How does the teacher determine criterion-related validity?***

- ▶ The teacher compares test results with some other independent test or evidence that is presumed to have high validity – this is called concurrent validity.
- ▶ In short, other means of assessment evidence are used to corroborate the test scores (triangulation).

▶ ***18. What is reliability and how does it relate to validity?***

- ▶ Reliability refers to consistency, both in terms of the internal pool of test items, and the external ‘pool’ of the application of the instrument from administration to administration. How consistent are the test items across the whole test pool when indexed against one individual? How consistent is the test from administration to administration. The key here is that there should be little variance.

- ▶ A well designed test should give a closely similar spectrum of results in a completely different school, for example, provided the material it is testing is taught in a similar manner, with a similar amount of coverage, and similar overall teaching quality.
- ▶ If there is a discrepancy in test results from such a scenario, there is an issue with the test's reliability.
- ▶ Reliability and validity are closely interlinked. If a test is unreliable, it is unlikely to be valid, so will almost certainly need further revision/improvement.
- ▶ If a test is shown to be valid, by meeting **1.)** a determination of the validity of *all* of the individual test items from our calculations of difficulty index and index of discrimination), and **2.)**, further, a determination of validity of the test instrument as a whole in regard to its content, criterion, and construct related evidence of validity, then we can state with some confidence that it should prove to be a *reliable* test.
- ▶ Establishing true reliability may take time, however, as it undergoes the real trial of extended use over time across many cohorts of students. For this reason, tests are frequently subject to almost continuous gradual revision/improvement or periodic adjustment, since weaknesses/deficiencies may be revealed that were not originally apparent or the particular objectives that need to be tested may change or mutate over time.

Action reflection/Self reflection

What you thought you understood and presented well

I feel I have presented the key part of this course, namely performance outcomes and the ABLE model effectively.

I also feel that I have successfully managed to bring together all the sections and material into a strong coherent order.

Improvements that you intend to do something about and how.

The jigsaw task proved particularly problematical, not just in terms of the actual group work but also in terms of the difficulty of compiling all the resultant group's works into one coherent document here in reasonable manner, which also took considerably longer than I anticipated.

In terms of the group work itself, I think I need to be more assertive in getting our group to do more planning first (regardless of the task). We were disadvantaged by starting off with a small group of four, then losing one member to make us yet smaller. Nevertheless, all of these difficulties are in fact merely representative of the difficulties one will encounter in real life, all of the time. One simply has to deal with the situation at hand and do the best job that you can in the given circumstances.