Welcome to topic three A where we're going to be looking at the principles and types of assessment.

Let's start with the etymology of the term assessment. It has a rather interesting and unusual origin. It originally meant in 1423 to fix the amount of a tax or a fine. And it literally came from the idea of sitting beside someone as they worked through their finances and receipts, etc., etc. This was precisely the role of the judge's assistant, whose job it was to help fix the amount of the fine or the tax by 1934. This had taken on the idea of to judge the value of a person or idea, etc.

Once the idea of assessment became ingrained in education, it took on the idea of the process of identifying, gathering and interpreting information about student learning. It was understood as any systematic method of obtaining information from tests or other sources. It is a process of drawing inferences from the material that was gathered about a student's performance. Ultimately evolving into the idea of the measure of a student's skills or knowledge in a specific subject area. Armstrong said it well in 1994, when she described it as the process of judging student behaviour or product in terms of some criteria, which may include objective tests and the use of writing scales, observation checklists, content analysis, interviews based on performances, discussions as well as written assignments.

So why do we assess? Well, assessment is really an integral part of instruction as it determines whether or not the goals of education are being met. Assessment affects decisions we make about grades, the placement of students in particular classes, their advancement in terms of the curriculum, instructional needs, even the way we design our curriculum, and in some cases it can even impact funding. Assessment inspires us to ask questions such as Are we teaching what we think we're teaching? Are the students learning what they are supposed to be learning? Is there a better way to teach the subject and therefore promote better learning?

What and how we assess is actually influenced by the curriculum ideologies we hold. We've looked at the curriculum ideologies in one of our earlier lectures. There's really four major curriculum ideologies that we often see in education. We have the scholar academic kind of ideology, which is really about inducting students into powerful subject knowledges. And the purpose of assessment in this particular ideological position is to rank students for a future career in an academic disciplinary field. So, we see in this kind of model grades that range from, you know, fail grade and then to pass right through to high distinctions, for example, or something like in A to E scale even because we want to see where the students act in relation to other students against the criteria that we’re set in the discipline. This contrasts with the student-centred ideology where the teacher's goal is to support students to achieve their full potential. Here we're not really interested in how students perform against others, But we want to understand the students performance in relation to their own capabilities. So, assessment becomes a tool to diagnose students abilities and capabilities where they're at right now, and we use that to inform future lesson planning to best support students learning needs. We then have the techno rationalist ideology or techno efficient kind of ideology. And here the goal is to produce workforce ready graduates. Importantly, this ideology, we certify people in terms of a set of competencies, and we do that so that a future employer or agent who they're going to interact with can see what competencies this person has attained, what skills that they've attained through the process of their education. This is the common form of assessment in various forms of vocational education, for example, but we will see it in curriculum areas where there are more skills focussed. Finally, we have the social Reconstructionist ideology where the goal of the curriculum is seen to transform students into critical and informed citizens. In this model assessment is often understood as measuring a student's progress with respect to certain perceived capacities and abilities. This particular ideology is also very interested in the values that students adopt and sometimes can be looking to see if they've adopted the set of values that are actually being advocated within a particular course. So, depending on the curriculum ideology that we personally adopt, this will change the way we think about assessment and why we're doing the assessment we're doing. And we can certainly see this across various education systems and various levels of education where the ideology shift. And of course, it's possible to be engaged with more than one ideology simultaneously. And so, we can see a focus, for example, in New South Wales, education on helping students achieve their full potential while simultaneously trying to induct them into the powerful subject knowledge, at the same time producing skills or competencies that are of a generic nature and useful for the workforce while still trying to develop into critical and capable citizens. So, in reality, all of these ideologies can often interweave and mix and assessment will change and transform according to where the particular ideological emphasis is placed.

Now, when we think about types of assessment, we could distinguish two major types. One is the idea of formal assessment and the other informal. When we think about assessment, we're often thinking about formal assessment that is an end of term exam or a particular assignment that has been set for which the sheet will be given a set of marks that are recorded. Informal assessment, on the other hand, is the kind of assessment that teachers are doing all of the time. Every time a student answers a question or engages with the task, the teacher is looking and thinking about whether or not the student has understood what they're teaching. This is a kind of informal assessment which is pretty much constant throughout the process of teaching and learning. And teachers will be adapting the way they're teaching or the content that they're teaching, depending on what they're seeing in terms of student response to the teaching methods that they're using.

We can also make a distinction between what we call formative and summative assessment. Formative assessment will be those kind of tasks that we position along the way in a unit, for example, that actually show us where the students are understanding or have the requisite skills and knowledge that we need them to have to be able to achieve the outcomes of the unit and will be giving them feedback along the way that will then help them improve their performance so that when we get to the end of the unit and where having the students complete the major task of the unit, whether that be an exam, for example, or whether that be some kind of a project or other assignment they've been doing throughout the course, that they'll be successful at that summative assessment task. So where is the formative assessment task feeds this information along the way about how students are progressing? The summative task tells us finally, whether they were able to achieve the outcomes that we set or not. While it is often, though not always the case that a summative assessment will be a formal assessment, formative assessments can be formal or informal. It really depends on whether we have set a specific task at a particular point in the unit or whether we're simply giving feedback along the way with whatever tasks the students happen to be doing.

Another way we can think about the types of assessment is to think about them as assessment of learning, assessment for learning or assessment as learning. Assessment of learning assist teachers in using evidence of student learning to assess achievement against outcomes and standards. Sometimes this is referred to as summative assessment. It usually occurs at defined key points during a unit of work or at the end of a unit term or semester and may be used to rank or grade students. The effectiveness of assessment of learning for grading or ranking depends on the validity and reliability of the assessment tasks themselves. We'll talk more about that later. Its effectiveness as an opportunity for learning depends on the nature and the quality of feedback. Assessment of learning is used to plan future learning goals and pathways for students. It provides evidence of achievement to the wider community, including parents, educators, the students themselves and outside groups, and provides a transparent interpretation across all audiences. Of course, standardised tests are often seen as a form of assessment of learning.

Assessment for learning involves teachers using evidence about students’ knowledge, understanding and skills to inform their teaching. Sometimes referred to as formative assessment, it usually occurs throughout the teaching and learning process to clarify student learning and understanding. Assessment for learning reflects a view of learning in which assessment helps students learn better rather than just achieve a better mark. It involves formal and informal assessment activities as part of learning and to inform the planning of future learning. It includes clear goals. The learning activity provides effective feedback that motivates the learner and can lead to improvement. Reflects the belief that all students can improve, encourages self-assessment and peer assessment as part of the regular classroom routines and involves teachers, students and parents reflecting on evidence. It should be inclusive of all learners. In this form of assessment, Assessment is working not only for the teacher to determine whether students have achieved outcomes, but also for the students themselves to see how they are tracking with regard to the learning intentions. And to help teachers work out what additional things they need to plan in order to help the students achieve those learning intentions or learning goals.

The final form of assessment in this triumvirate is assessment as learning, and this is the one that most often confuses people. Assessment as learning occurs when students or their own assesses students, monitor their own learning, ask questions and use a range of strategies to decide what they know and can do and how to use assessment for new learning. Assessment as learning can also be where we've set tasks that themselves drive the understandings that students need to develop. So, by setting a particular task, we know that they're going to have to develop certain capacities that are exactly the learning goals that we're aiming to support the students to achieve. Assessment as learning encourages students to take responsibility for their own learning. It requires students to ask questions about their learning, involves teachers and students creating learning goals to encourage growth and development, provides ways for students to use formal and informal feedback and self-assessment to help them understand the next steps in their learning and encourages peer assessment, self-assessment and reflection.

When we're designing assessment tasks, we need to keep in mind two important concepts. What we call validity in assessment and reliability. Let's look at validity first. The validity of an assessment tool is the extent to which it measures what it was designed to measure without contamination from other characteristics. For example, a test of reading comprehension should not require mathematical ability. Likewise, a test of mathematical capability shouldn't be overly dependent on a student's reading comprehension level. During the era of the so-called White Australia policy, more correctly called the Immigration Restriction Act, people from non-English speaking countries were often given tests in a language other than English, but not a language that was their own, or they were given tests in English that tested for cultural understandings that they wouldn't possibly have because even if they spoke English well, they were being tested on ideas and concepts in Australian culture that they simply wouldn't have been exposed to. In each of these cases, the tests that were being given were hardly valid for the purposes for which they were arguably being presented. Another important example for the idea of validity is that we assess actual performance. So, for example, a valid driving test should include a practical driving component and not just a theoretical test of the rules of driving. When you come to your Newcastle teacher performance assessment, one of the important elements of our particular model here at Newcastle is to incorporate an in-situ assessment of your pedagogy, or at least an evaluation of your pedagogy done by a university supervisor using the New South Wales Quality teaching model. Other forms of the teacher performance assessment that are out there in practice by the universities, often ask students simply to film a short segment of their class and to provide that as evidence of their teaching capacity. But of course, in such a model, that short ten-minute video might have been the best 10 minutes of the class and the rest of the class might not have worked very well at all. And so there's a question about whether or not that's a valid form of assessment of one's classroom teaching capability. Likewise, if we simply asked you to write an essay about how or report on how well you did in the classroom, and we accepted that as evidence of your capacity to teach, then again, it wouldn’t be a very valid test of your actual performative capability as a teacher. So, for a test to be valid, we must be assessing what we say, we're assessing. And if the thing that we're assessing is a skill of some kind, then we would expect to see some kind of performative task involved in the assessment, not simply a theoretical writing task or report.

Reliability is the second component of a good assessment. Reliability refers to whether an assessment instrument gives the same results each time it is used in the same setting with the same type of subjects. Reliability essentially means consistent or dependable results. Assessments are reliable if they produce comparable outcomes with consistent standards over time between different learners and examiners. In other words, if two different teachers were using the same rubric to evaluate the same assessment task, they should come up with roughly the same result. The closer their ratings are to each other, the greater interrater reliability that particular instrument would have. If different teachers given the same assessment task, using the same instrument, produce radically different assessments of the student's work. Then we would question the reliability of the instrument that's being used to make the assessment. Or of the teacher’s judgement themselves. In school, the process of gaining interrater reliability is referred to as consistency of teacher judgement. And this happens through a process we call moderation or where teachers sit down together and look at their evaluations of a particular group of students work and see if they've come up with the same kind of ratings. It's important to realise that a test could be reliable but not valid. For example, if you have a scale that you measure your weight on every morning and that scale happens to be off by five kilograms, it reads your weight every day with an excess, let's say of five kilos. The scale is reliable because it consistently reports the same weight every day, but it is not valid because it adds an extra five kilos to your true weight. In this case, we could say that the scale is reliable, but the rating that it's given us is invalid. Again, if we related to something like the NTPA, if we simply asked you to describe the effectiveness of your teaching without actually ever looking for evidence of what you did in the classroom, then we may well get a highly reliable result. The writing piece itself may be something that many NTPA examiners could all examine and give you identical marks for the work you produce, but it wouldn't really be telling us about what your capabilities are in the classroom. To do that, we need some evidence of what you've actually done in that setting. This is why the TPA requires the collection of evidence, as well as your annotations on, that explain how this evidence you've collected represents your achievement of certain professional teaching standards. Doing this, we hope, creates a more valid and reliable type of assessment of your teaching performance and impact on student learning.

One way to think about the design of assessment that works quite well is what we call smart assessment or smart goals. Here, the mnemonic SMART is used to ensure that you've got all the key features you need to make your assessment work effectively. And this, of course, works particularly strongly with outcomes based education. So using the smart concept, the task we set must be specific, it must be measurable, it must be achievable, it must be relevant to the topic and it must be time specific. Well, we have each of those components. We have designed a smart assessment that actually can work effectively in the classroom, as I say, particularly in terms of outcomes based education.

What the research shows is that assessment works best when it provides diagnostic feedback, when it helps educators set standards, evaluate progress, relate to a student's progress specifically, and motivates performance.

One important thing to remember when designing your assessments is to look at the verbs that are being used in your syllabus outcomes. These verbs give clues about the kind of task that you should be asking students to do. The verbs themselves often align well with Bloom's taxonomy, which could be used as a basis for assessment. So, for example, if we think about whether the student can define duplicate lists, memorise, recall, repeat or state something, we're really asking them to engage in remembering and we're asking ourselves really the question, can students recall or remember the information? If we find verbs like classify, describe, discuss, explain, identify, locate, recognise, report, select, translate or paraphrase. We're really asking whether the students can explain ideas or concepts, so we're seeking to determine their level of understanding. When we see words like choose, demonstrate, dramatise, employ, illustrate, interpret, operate, schedule, sketch, solve, use or write, We're seeking to find out if students can use information in a new way. In other words, are they at the level of applying the information that they've learnt? Where we get students to appraise, compare, contrast, criticise, differentiate, discriminate, distinguish, examine, experiment, question or test. We're seeking to find out if the students could distinguish between different parts of a topic or problem or idea or concept or skill, in which case we're asking to see if they've achieved the level of analysis or the ability to analyse. When we ask students to appraise, argue, defend, judge, select, support, value or evaluate. We're seeking to find out if students could justify a stand that they're taking or decision. And so we're really working at Bloom's level of evaluating. Finally, if we're asking students to assemble, construct, create, design, develop, formulate, or write, we're asking students can they create a new product or point of view? Or perhaps even develop a performance of some kind. At this level, we're looking at Bloom's creative stage. The tasks we set should match whatever the verb we are using. And so if the syllabus is asking us to get students to describe then the kind of task we set would really be a task in which students do some describing if the task is asking or the verb is asking you to appraise, then the task we design should have students appraising or distinguishing between different parts.

Finally, it would be remiss of us to be discussing assessment without thinking about, once again the concept of backward design. In backward design, we're beginning really with assessment. Of course, we start by determining what we want the students to learn about or learn to be able to do. But then we design a task that will allow us to collect evidence about whether students can in fact demonstrate their knowledge or ability to do the things that we've set as the learning intentions. From there, of course, we plan our learning activities to support them to be successful at that assessment task. It's also become popular to think about this as beginning with the end in mind. So in other words, we're really thinking about what is it we want students to do at the end of the unit. And in that regard, how that task itself aligns with the outcomes. And from there, we plan our teaching learning activities.

Let me finish by asking you what assessment of learning, Assessment for learning and assessment as learning look like in your subject area.

Thanks for listening.