Episode 7: Dr. Josh Weller

**KL:** Katie Linder
**JW:** Josh Weller

**KL:** You’re listening to *Research in Action*: episode seven.

[intro music]

# Segment 1:

**KL:** Welcome to *Research in Action*, a weekly podcast where you can hear about topics and issues related to research in higher education from experts across a range of disciplines. I’m your host, Dr. Katie Linder, director of research at Oregon State University Ecampus.
On this episode, I’m joined by Dr. Joshua Weller, an assistant professor of psychology at Oregon State University. Dr. Weller received his Ph.D. in Psychology from the University of Iowa.  His research broadly focuses on how affective and cognitive processes contribute to decision-making and risk perceptions and, more particularly, on the development of psychological scales to quantify individual differences in risk taking tendencies and decision-making competence. His research has been funded by the National Science Foundation, the American Automobile Association Foundation, and the National Institute of Drug Abuse. Dr. Weller teaches courses on Judgment and Decision Making, Personality, and Psychometrics. Welcome to the podcast Katie.

**JW**: Thank you Katie, I’m glad to be here.

**KL:** Good! So I feel like I should tell our audience that you and I are colleagues here at Oregon State University and I’m happy to admit that our first meeting was because I knew nothing about something that you do very well. And I had been talking to a consultant that I work with, we had been working on putting together a survey instrument and she said, “Well do you have a psychometrician on your team?” And I said, “No.” And I very quickly scrambled, went on Wikipedia, looked up psychometrics, Googled psychometrics in Oregon State, and found you, and invited you to coffee. And we had a chance to chat. So, I became really interested in kind of thinking about what is psychometrics and was really pleased that you could join me on the podcast. So, for our listeners who may not be familiar with it, let’s start there. What is psychometrics?

**JW:** Ok, so briefly and really kind of at the broadest level psychometrics is the scientific study of the attributes of tests, of psychological measures. What we’re trying to do is assess in the broadest terms, the quality of a psychological test to make sure that it’s reliable; it measures what it’s supposed to measure. And that has really important implications for the use of tests kind of at the end user stage and test administrators, the people taking the test, interpreting the scores and knowing what that score means is very, very important.

**KL:** So I’m wondering if we can talk a little bit about, because as I got into kind of thinking more about survey design and reading more about it, a couple of things that came up immediately were things like reliability and validity. So, when we think about psychometrics are those kinds of key areas of psychometrics? And, what exactly are those things?

**JW:** Sure, sure. Yes, they are both very, very important issues with respect to tests. The concept of reliability is the matter of taking the same test and getting similar scores over repeated measures, over repeated time periods. The more stability that you have over time, the more reliable the test is. So if you’re taking a test at time one, an IQ test, and you have a highly reliable IQ test, you take it five years later your score shouldn’t be that much different or the correlation should be very high between scores at time one and time two. So what we’re looking for is the more, the lower test retest correlation when we think about reliability in that aspect, the greater the correlation the more the reliability. The lower the correlation between two testing assessments means that there’s more error, there’s more measurement error there. So that makes it a less reliable instrument.

Now in terms of, there’s other ways to measure reliability and it doesn’t require you to take tests at time one and time two. One of the most common ways to do it is to take items of a scale and look at the internal consistency of those items. So when you’re doing a personality test or an IQ test you run through a bunch of items: I agree with this or this is the right answer or this is the wrong answer. And what a psychometrician will do or a researcher will do is look at the correlations between those items and see how well they hang together. If they are hanging together strongly or to an acceptable level we can say that they may be caused by some type of overarching construct; this latent variable that we deal with in psychological tests because we can’t really, we don’t directly see things like extraversion, IQ. We can only estimate it or approximate it by observable behaviors. And this is kind of what we’re trying to do with this term of internal consistency; trying to figure out do these behaviors cluster together in a meaningful way. So that’s one of the ways that we, another way that we look at reliability.

Turning to validity, you can have reliable measures and they might not be valid. Just like as if you’re shooting at the dart board and you’re trying to hit the bullseye and you hit the one fifteen times in the outer ring. Yeah, you have good groupings, but you’re not hitting your target. So validity is a concept of whether or not the test is measuring what it says it’s measuring. And to achieve validity it’s a continual process. This idea of construct validity; does the construct measure, does this hypothetical construct measure in the test that we have measure what it’s supposed to be? And that goes, we go through lots of different processes by which we try to establish validity knowing that validity never, the quest for construct validity never stops. It’s a journey rather than a destination. And with respect to validity it’s not necessarily a property of the test itself, but it’s a property of the context. So a test could be very valid for one reason, but it could be very invalid for another reason. Just like thinking about tests as tools. A hammer is a very good tool for knocking a nail into the wall. It might be less of a good tool to try to open up a can.

**KL:** So one of the other words you used kind of frequently in those explanations was construct.

**JW:** Yes.

**KL:** But that also seems like a very foundational component when we’re thinking about psychometrics. Can you talk a little bit more about what you mean by construct and maybe give a couple of examples?

**JW:** Oh yeah, sure. So at the heart of psychological tests why do we give these test like extraversion or big 5 personality tests – extraversion, agreeableness, neuroticism, openness, conscientiousness. You can’t really see those things in the open; you can’t see extraversion, you can’t see openness. You can only see glimpses of observable behaviors. So what we’re trying to do when we create a psychological test like a personality measure is we’re trying to create a sample, a representative sample of behaviors that we predict or theorize go together. And that’s what we mean by a construct. It’s a theoretical entity that we can’t observably see, but we can infer it from representative behaviors. And if those representative behaviors are correlated with one another and are related with one another in a particularly meaningful fashion and a consistent fashion, we can say that according to this latent variable theory, this construct theory, that what’s causing those correlations is because of that latent variable. It’s because of variability in that latent variable between you and me say in extraversion or in openness. So, in other words, if we, you know, if you’re more extraverted than I am, right? Your scores on questions like I like to go to parties or, you know, I’m very talkative, energetic would be higher than my scores on those observable behaviors. But what’s causing that? And we would say it’s because you’re more extraverted than me.

**KL:** So it sounds like some of the work of working with constructs is to some degree labelling constructs correctly. But it sounds pretty complicated. Is that a correct way to think about constructs or is it something different?

**JW:** Yes. There’s an amazing amount of subjectivity that often happens with this, so you have to be very careful. We talk about one form of validity being face validity. Do the items look like what you’re saying that it’s measuring? And that’s not sufficient enough, right. It can look like a duck and it can quack like a duck, but it might not necessarily be a duck or you might be missing important parts of the duck. You know? You need to be able to fully characterize it. So when you’re naming things, when you’re naming constructs, when you’re naming these theoretical entities you have to take a look at the literature. You have to look at different perspectives and understand from different angles what people have thought about when you’re trying to quantify something like being a risk taker or something to that affect. You have to think about, you know, what causes somebody to be a risk taker? What are the types of risks that people do? Those kinds of things. And when we get into kind of some types of mathematical equations, like factor analysis, and it spits out different dimensions it becomes a challenge and always a challenge for a researcher to accurately label the dimensions and make sure that they’re interpretable. And you want to make certain that these things are interpretable and that people can agree on them. Say, this really seems like this is what it is and this is what this construct measures. And then you sort of subsequently, you kind of follow that up with other studies that support that assertion.

**KL:** That seems to be the objective part about creating the objective measurement, is something that people can agree on, they can look at it and that it makes sense to them. To what degree are people who might be the audience for the instrument or who would be the people kind of engaging in the measurement. Are they taken into account at all when you think about labelling constructs or even just the context or the, I think about some of these measurements like you were talking about with extraversion or other kinds of things, that depending on the age of the audience you’re thinking about, for example, that could really impact the kinds of questions that you ask them or the ways that you might think about scoring an instrument. So I’m wondering if you can talk a little bit about that; where do those factors fit in?

**JW:** Yeah, all of those come into play at certain levels or another. So, at the very basic level when you’re constructing questions you want to follow, you know, really best practices about the wording of questions. You don’t want to include things that are slangy or have a lot of industry jargon in it. You don’t want to have word levels, you know, that are higher than, you know, what the average person is. Make it easy; keep it simple, keep the questions straightforward so there’s no ambiguity in how people respond.

As far as over the lifespan it becomes a challenge, especially if you’re trying to look at developmental trends in traits or developmental trends in constructs because obviously there’s clear differences in reading level at, say, age 8 versus age 15 versus 25. So what we can try to do is, you know, we can modify questions and then try to, you know, correlate those as people kind of move on, as we develop questions try to correlate them with past instruments to say, “Ok this is an approximation of the last one” in a very simple situation. But you always have to consider things like, for instance in risk taking, if you are, a lot of popular risk taking measures or risk propensity have items like, you know, going bungee jumping and sky-diving. For instance, if you try to give that to people who are octogenarians they might respond very low. You might get a lot of floor effects because they just don’t have a real interest in doing it. But it might fail to see some of the other types of risk that they face every day. So you have to really be, keep in mind your audience when constructing a test, when administering a test.

And this is why this issue of face validity that I was saying, you know, it looks like a duck and quacks like a duck, you can’t just apply it to any particular population willy nilly. And this goes into validity, right? They’re valid for purposes, they’re valid for particular situations, but not all. So it’s one of those things that you always have to keep in mind about, you know, when you’re administering the test because tests have consequences and they’re good, they can be good consequences and they can be negative consequences. And if a test is administered in an inappropriate way or interpreted in an inappropriate way, it could lead to denial of resources at its worst kind of, the worst type of consequence. So that’s not something that we want to do. So we always want to be cautious with the interpretations, we want to when administering the tests have people that are trained on the instruments, that know how to interpret things. So when they do get the results they’re not over-extending those results. They’re not saying that it means something that it doesn’t and so on.

**KL:** This is a really wonderful, concrete introduction to what psychometrics is. After a brief break we’re going to come back and hear a little bit more from Josh about how does one train to do work in psychometrics. What is kind of his background in that as well and also a little bit more detail about his own research.

[music]

# Segment 2:

**KL:** So from our beginning conversation Josh it sounds like psychometrics can be very nuanced, pretty complicated. I’m wondering what kind of led you into using this as part of your research work. What led you to this part of your field?

**JW:** That’s a good question. So I’d always been and I think one of the things that always drew me to psychology when I was younger, when I first thought about it being, you know, a career. Even as young as seventh, eighth grade I’d thumb through my father’s psychology textbooks and I’d always see tests and measurements of, you know, what personality are you and things like that. I’d do them and give them to my friends and try to score them and whatnot. I always had a real affinity towards statistics and actual measurement of different types of qualities. So from there, you know, when I actually went to graduate school for psychology I went into a personality and social psychiatry program and started to, I had kind of had, initially had this kind of Pollyannaish feel of what grad school was all about. And it was like, oh I’ll be a professor and I’ll be, you know, teaching and imparting knowledge on everything. And then it’s like, damn there’s just statistics everywhere! So I said, “Wow, I’ve become a statistician” and I embraced that.

One of my, you know, greatest influences on this was with not necessarily my advisor, my primary advisor, but a secondary advisor, David Watson, who has done lots of work in psychometrics and affect and personality. And I’ve learned a lot from his courses and, you know, working with him directly about factor analytic techniques and different techniques about how to appraise personality, how to appraise psychological tests in terms of reliability and validity. And then from there I’ve just started to apply it to the knowledge or this interest in individual differences. Looking at how individual differences in personality may relate to risk behaviors in individual differences in how we respond rationally to different types of problems. How does that relate to health outcomes, social outcomes, and so on? So I’ve kind of, it’s been an evolving process from, you know, all the way back in seventh grade to present day.

And today, you know, we’re working on some, you know, more scale related types of inquires where we’re looking at understanding, you know, the structure of risk taking and, you know, how many, what kind of domains of risk taking are the most important or, you know, what is added, this one’s attitude towards certainty and can we make self-report measures of that to understand are people more fearful of the unknown? So, what we’re trying to do is quantify that and see where those behaviors, where those individual differences may relate to psychosocial outcomes, not only just interpersonal, political attitudes, but also, you know, even psychological disorders.

**KL:** Well and recently you’ve had some kind of interesting funding for your research coming out of the American Automobile Association Foundation.

**JW:** Oh yeah.

**KL:** And the National Institute of Drug Abuse. What is some of the work that, it sounds like very concrete work, very applied. What are some of the projects that you have associated with that?

**JW:** Yeah, that project with the American Automobile Association Foundation for Traffic Safety was actually my first grant back in like 2007, 2008. And they had approached us about trying to understand distracted driving behavior.

**KL:** Interesting.

**JW:** In young adults. So what we did with that project, it was kind of a two-stage event because we didn’t really know what young people thought about distracted driving at the time. It was fairly new and it was a fairly new concern as a social issue. So we did some preliminary focus interviews to try to understand what teenagers were thinking about distracted driving. Did they see it as a problem? We found some very interesting results that came from there that were somewhat staggering. That some teens felt that driving got in the way of their phone calls rather than phone calls got in the way of their driving.

**KL:** Ok.

**JW:** So yeah. People found, also reported in these preliminary interviews a very deep attachment to their cell phones. And I wouldn’t quite call it an addiction to a cell phone because that kind of implies some physiological process going on. But there is a certain attachment that I feel lost without my phone. I have a moment of panic that rushes over me if I leave my house and my phone is laying on the bed. I’d rather lose my wallet than my phone. And these kind of ideas were really echoed. What we did from those interviews and through looking at the research we constructed some scales and some measures of risk appreciation about cell phone use while driving, as well as cell phone attachment. How attached you were to cell phones, to your cell phone.

And we found that in a nationally representative sample of people that people who had higher risk perceptions about cell phone use while driving, like they felt that it was dangerous, you know, they felt it was dangerous, they felt that it affected somebody almost as much as driving with a little bit too much to drink, maybe just under blood alcohol level, that it affects everybody on the road not just the driver. People who endorsed those types of questions actually didn’t use their phone as much. They weren’t as likely to text while driving, to talk while driving. Conversely, the people who felt very attached to their phones, as you might imagine, those were the people that, they were texting more. They were texting a lot more and they were talking more, as well. And this seemed to be sort of an age-related phenomenon that it was especially prominent in younger users rather than older users, especially for SMS. So we found some really neat insights on that and we’re kind of building off that. We’ve collected some data in an Italian sample with maybe almost 1,000 people from an Italy community, a sample about risk appreciations of cell phone and distracted driving use in Italy at this point. And we’re writing that up with one of my colleagues from Italy and some of my research assistants here at OSU.

**KL:** That is fascinating and, you know, it’s so interesting to think about you may start out thinking you’re going to get something and then something completely different comes out of the data and you can kind of follow that pathway. Part of what this raises for me is a question about psychometrics and correlation versus causation. Can you talk a little bit about that? I mean I think that, I’m going to ask it very broadly, you know. How do you see that in your work in terms of what people might think these kinds of measures can tell us?

**JW:** So this is always the, you know, kind of the almost overlooked issue even though so many people, not necessarily in psychological science, but in kind of the broader community that I think get it wrong even though they can say correlation doesn’t equal causation until they’re blue in the face. You see popular science reports, you know, this leads to this and without singling out any particular writers or any particular studies, you can go back and go, “Wait, this was a correlational design.” Right? There was no random assignment. So when we talk about correlation, and this is most of what we do with psychometrically based tests is correlation method. We’re trying to associate a score with some outcome, some predictive outcome, and we’re trying to use that score to say what’s the likelihood that the person with score X is going to do this at a certain level. It’s not a yes, no; it’s a propensity, it’s a probability in a way, we can think about it.

So it’s not really, it’s not causation. One doesn’t cause another because when you have correlations between things you can have very high correlations between things and they’re completely spurious. You can have correlations between two variables and it might be caused by a third variable, right? I use the example in my class and you know some people may have always heard this before. You know, in large US cities there’s a small correlation between murder rates and ice cream consumption. It doesn’t mean that, you know, eating ice cream makes you want to kill people. That’s a causal link, no. So why might that be the case? There’s lots of third variables that could explain that correlation. Similarly with correlation-causation one of the reasons why you can’t infer that is because there might be reverse causality.

So we say, for instance, being disagreeable might relate to higher levels of drug use. So is there, where is the pattern of causality? Where are we establishing temporal precedence? We can’t do that when we’re not manipulating variables. If we manipulate variables we can start to, while controlling for other things, we can say more certainly that this causes that. But with correlational designs we tend to just go, we tend to measure the two independent variables. And as such we’re unable to infer that kind of causal inference. Unfortunately when you see popular press things with correlational designs that often gets overlooked for the sake of sensationalism. Oh, this causes depression or this causes that without having anything that’s longitudinal or anything that’s been manipulated or anything that’s even smacking of something that is, you know, an experimental design, so.

**KL:** Absolutely. I think we’ve all seen studies like that, that are sensationalizing and really pushing that causation piece. In the first segment we had briefly talked about validation and I’m wondering to what degree you engage in the validation of measures or instruments in your own research and what that process might involve because it can be kind of complicated.

**JW:** Yeah, yeah. So when we think about validation we think about it in a lot of different aspects. And some of the basic aspects are do the items, at kind of this face validity question are the items that you’re including in a psychological test, do they represent the theory that you’re trying to say? Is it theoretically consistent with, you know, are those items theoretically consistent with your construct?

The second part is do these items go together? Do they make up the construct? Do they make up this, you know, thing you can’t see? This imaginary thing that you can’t directly see, but you can infer. But then you also want to start looking at, you know, whether or not these things correlate with other types of scales that might be related and not correlate with scales that are supposed to be not related to the scale. So you’re looking at kind of convergence and discriminant types of validity are what we would call it. We want to make sure that our items, you know, are traits. There’s lots and lots of traits out there where if we create a new scale we want to show why the world needs our scale, right? So we want to show that it converges, it’s similar to some traits that you would expect it to be related to, but not too close because then why do you need it? But you also want to say that it’s distinguished from something else. And distinguished from another alternate hypothesis or an alternate hypothesis.

From there you also want to consider things like predictive validity. You want to see things like, does this predict what you have theoretically envisioned it to do. You know, some, I work with a trait called honesty-humility and you would expect people who are low in honesty-humility to have higher rates of absenteeism at work, you know, workplace dishonesty, greater levels of sexual harassment, those types of things. And you do find that and that would be an example of honesty-humility having predictive validity in this case. So what we would, similarly with risk taking. If you score high on a risk taking test and you’re presented with a risky behavior, are you going to take the risk versus, you know, an uncertain choice versus a safer choice that has maybe a lower outcome, but it is more certain.

**KL:** So it sounds like when we’re thinking about building instruments, building measurements, making sure they’re reliable, making sure they’re valid, and you were just pointing out like there’s a lot of theory involved. There’s a lot of kind of thinking about how constructs relate to each other. Is this really measuring what it’s supposed to measure? For people who are just starting out in this area, they may feel and I think this is the case with a lot of methods. They may feel incredibly abstract and if you have not applied it, it can feel overwhelming. Do you have any resources or kind of go-to things that you could recommend for people who are just starting out in this area and who are starting to kind of just wrap their brains around the concept of what psychometrics is?

**JW:** Yeah. So as far as I think first of all learning a base of statistics is very important. Even at kind of the rudimentary level there’s some good websites that kind of teach things like SPSS. A guy named Andy Field does a really kind of funny, comical type website on learning SPSS. It’s called Statistics in Hell or something like that.

**KL:** Well we will definitely link to that in the show notes for people, yeah.

**JW:** It’s pretty funny because it has like ok there’s one ring of hell, then a second ring, then the more complex things are on the third ring. So I always find that kind of funny and sometimes I go back to that if it’s something I haven’t done in a while as far as just basic type work. He also has some books on, you know, using SPSS to do statistics. As far as kind of just learning about it there are a lot of good introductory textbooks, the textbook by Fer and Bacher, which I use in my class is a good resource. The Anastasi-Urbina psychological testing book is another one that’s more based on psychological appraisal, uses of psychological tests, how to deal with scores. I actually will use that in my graduate level course. So those are some kind of academic type things.

As far as the kind of coming from non-academics there are a lot of classes about I believe the MOOCs I believe they’re called.

**KL:** Yes, the massive open online courses.

**JW:** Yes. So I’m guessing that there are several of those available that somebody could get into and sort of learn about some of the basics. I mean at the basic you really need to know kind of the concepts of reliability and validity and factor analysis. And then kind of go from there at the simple, but even more, I wouldn’t say even more important, but equally as important is learning the subtleties of scale construction. This psychometrics is a garbage in garbage out enterprise. Like many things and if you don’t have clearly constructed items, if you have very ambiguous things that people are responding in too many different ways, you’re asking questions that make people defensive, overly defensive. These are all going to kind of creep up on you and moreover kind of thinking about the theory, thinking about scale construction as a theory taking into consideration all kinds of different points of view while writing those items, while writing those questions. And then doing some other techniques, like you know I’ve talked about factor analysis a couple of times to kind of let the data analyst sort it out. Say which items, you know, represent this dimension and which ones don’t. So kind of moving from there.

As far as those resources go I don’t know one, as far as like how to write good survey items off the top of my head I’ve kind of went to some, went to articles that I’ve picked up along the way. And I can certainly share them with you as links to, you know, for the podcast and for the listeners.

**KL:** Absolutely, we will link to those in the show notes. The other resource I would add as well is there are several kind of statistics modules on Khan Academy, which is something that I have turned to to get a pretty basic idea of what are some of these things, what are kind of the basic mathematical principles that are behind them. So we’ll go ahead and link to those as well.

I want to thank you so much Josh for sharing your expertise and your knowledge on the episode with me today.

**JW:** Yeah, you’re welcome.

[music]

**KL:** Show notes with information regarding topics discussed in each episode, as well as the transcript for each episode, can be found at the *Research in Action* website at [ecampus.oregonstate.edu/podcast](http://www.ecampus.oregonstate.edu/podcast%22%20%5Ct%20%22_blank).

There are several ways to connect with the *Research in Action* podcast. Visit the website to post a comment about a specific episode, suggest a future guest, or ask a question that could be featured in a future episode. Email us at riapodcast@oregonstate.edu. You can also offer feedback about *Research in Action* episodes or share research-related resources by contacting the *Research in Action* podcast via Twitter @RIA\_podcast or by using the hashtag #RIA\_podcast.  Finally, you can call the *Research in Action* voicemail line at 541-737-1111 to ask a question or leave a comment. If you listen to the podcast via iTunes, please consider leaving us a review.

The *Research in Action* podcast is a resource funded by Oregon State University Ecampus – ranked one of the nation’s best providers of online education with more than 40 degree programs and nearly 1,000 classes online. Learn more about Ecampus by visiting ecampus.oregonstate.edu. This podcast is produced by the phenomenal Ecampus Multimedia team.

# Bonus Clip:

**KL:** In this bonus clip for episode seven of *Research in Action* Dr. Josh Weller shares how he uses statistics in his hobby of fantasy baseball. Take a listen.

Josh, I understand that you’re a fantasy baseball fan and that you kind of use some of your statistics skills and knowledge to dabble in that area. I should freely admit that what I know about fantasy baseball and baseball and statistics is totally related to *Moneyball* and kind of the popularizing of that, you know, through the film and through the book, which we will link to those things if people aren’t familiar with them. But can you tell me a little bit more about this area and how you apply your statistical knowledge there?

**JW:** So, not only do I apply my statistical knowledge, but I try to apply decision making knowledge as well because this is kind of an interactive game.

**KL:** Oh that is fascinating.

**JW:** So you play with other managers, other people that are equally as crazy as I am about baseball. This really started when I was young and the idea of what was called rotisserie baseball just opened up. And it was mainly, when it started, was people, you know, looking through and thumbing through the box scores of newspapers and, you know, marking off people who got runs and you basically selected a team out of a player pool of, you know, Major League Baseball players. And how they did in real life depended on how, you know, impacted how you did in your league or on your team.

**KL:** So, just to kind of clarify because I’m less familiar with this, and maybe our listeners are as well. So you, in fantasy baseball you construct a kind of an imaginary team of players from a range of different teams. Is that correct?

**JW:** Yes, yes.

**KL:** Ok. So you can kind of pick strengths of various people and then based on how they do in real life it impacts how things happen in your team.

**JW:** Exactly. And typically, you know, there’s lots of different formats that go on. One that I like to play has six categories of statistics for hitting and six categories of statistics for pitching. And then you get twenty or fifteen hitters and ten batters that you pick from a bunch of, you know, pick from a player pool. We typically eventually go into what we call auction format where we have an auction and we bid on players with money and with a set amount of pool. Say, ok, you can have any player you want, but you have to name your price basically. So there’s a lot of valuation that goes on with that and kind of assessing players’ talents, upside, downside risk, those types of concerns. What the player did last year, what they’re expected to do this year.

So as far as statistics go it becomes really important to try to accurately value or try to prognosticate the future. One insight that’s really come up in the last ten, fifteen years, maybe even longer, probably in the last ten, fifteen years, since the *Moneyball* kind of concept and even a little bit before then was that the traditional statistics of baseball, things like batting average, wasn’t necessarily a great indicator of performance. It wasn’t the best indicator of performance. If you looked at baseball cards back in the 70s you’d get, you know, these counting statistics of home runs and runs and stolen bases and the batting average. And one thing that we neglected to know with batting average, it only tells you if you get on base with a hit. So one thing that you want to as a baseball player or as a real baseball team, you want to be on base regardless of how you get there. So these sabermetricians, people who are, you know, dedicate their spare time or even lives to kind of coming up with metrics of performance that try to incorporate things that we might not have seen that actually might be better indicators than kind of the veneer that we are given as just kind of cursory types of information. Right? Kind of the basic stuff. So the looking underneath the surface to see if there’s pockets of value or pockets of, you know, you’re seeing improvement over years or you’re seeing declines over years that are being masked by some of these less informative statistics.

**KL:** To some degree it seems a little bit like a treasure hunt. You’re looking for the thing that’s kind, you know, what is that thing that’s going to break it open and you’re going to realize that’s actually not the real measure. This other thing is telling us what’s really happening.

**JW:** Yeah, yeah, for sure. And there’s a constant, you know, development and understanding of this. And this is a fairly new field and, in fact, it’s gotten to a point where even Major League Baseball teams have hired sabermetricians and mathematicians and it’s kind of moving to a new model of, you know, of understanding, scouting, those types of things that have kind of revolutionized in a way how, you know, how players are evaluated, how contracts are given out and so on. So a lot of it started in a way, you know, I wouldn’t say a lot of it started because of rotisserie baseball, but by these sabermetricians – Bill James is one of them, Rodney Chandler was another – who are these kind of older pioneers of this game that really started looking at data and really trying to make sense of very large pockets of data to try to predict future behavior.

**KL:** Now one of the things I think I remember reading about sabermetricians and I think it was actually in *Moneyball* was the concept of like it started with people really doing it as a side gig. Like it wasn’t their full-time job, it was like for fun. Like they were just kind of playing around with the data and trying to find some stuff. And it ended up being for some of them full-time work once they kind of got into it. And I think this is something that may also have been seen in other sports as well. That as the kind of mathematics behind the sport has become professionalized, people have kind of made their own gig in terms of being tied to a sports team or something like that. I’m wondering, you know, to what degree are there other areas, other than fantasy sports, where psychometricians or other kinds of measurement specialists are doing this kind of dabbling. Like, can you think of other areas because this seems to be such a huge one where we see people kind of playing on the side with these research skills and these mathematics skills?

**JW:** Well, I think anything that you’re trying to detect some type of trends or behavioral trends, you know, I don’t know if I’m risk tolerant enough to do stock market investing, but I would say that some of those tools are similar. At least as far as thinking about forecasts and, you know, thinking about what stock’s going to go up and trying to take into consideration different types of information, and trying to quantify things in a way.

**KL:** That predictive modeling.

**JW:** Yeah, exactly. So I could see that happening. One thing to always known, at least with rotisserie baseball or fantasy baseball, is with these predictive accuracies and even in personality traits and most psychological tests you’re only predicting with 70 to 80% accuracy. Which, you know, it’s not perfect; we’re predicting tendencies here and it’s not a direct science, it’s not always going to happen the way you do it because unforeseen things happen or, you know, using psychometrics an error happened.

**KL:** So I’m going to completely betray my sports ignorance here to ask is there a season for fantasy baseball? You know, when does this take place? Because it’s tied to real information, is there a particular time of year where this is really happen?

**JW:** Oh, you’re going to laugh because it’s a yearlong enterprise.

**KL:** Ok. Well tell me more. I’m really interested.

**JW:** So my league that I’m in, or one of the leagues that I’m in, is what we call a dynasty league where you’ve picked players and then you keep them for infinite. We have a very deep minor league bench where we are scouting players that are in single A, double A.

**KL:** That seems very involved.

**JW:** It is. It’s a sickness, but it’s fun. It’s very fun. You’ll see me scouring through, you know, scouting reports all the time and looking at, you know, statistics and looking at players’ performance relative to the league average in the Minor Leagues and scouting notes and so on. So there’s usually a lull after the World Series until about December. People aren’t really that active because football’s still in kind of their realm. But as soon as the Super Bowl’s over, even during January, people are starting to trade their players. They’re starting to read things and it’s start to ramp up again. Then you keep going and going until September. The games start in April and they go on until the beginning of October. You don’t usually have it in the playoffs for the format that I play. Now with these daily fantasy sports they play it, you know, whenever. But, you know, it’s a marathon. It’s a slow journeying marathon that takes a lot of commitment to do.

**KL:** Absolutely.

**JW:** This is for the very committed.

**KL:** Absolutely. So you mentioned that you play in a league.

**JW:** Yeah.

**KL:** How many folks are generally in a league?

**JW:** Well typical leagues are about 10 to 12, but the league that I was just talking to I have 20 players. So it’s a large, large group of people from all over the United States. And yeah, it’s pretty, yeah.

**KL:** Well it sounds like a lot of fun. Thank you so much for broadening my horizons in terms of thinking about how statistics can be related to these fun endeavors that you do on the side.

You’ve just heard a bonus clip for episode seven of *Research in Action* with Dr. Josh Weller sharing how he uses statistics in his hobby of fantasy baseball. Thanks for listening.

[music]

# “Research in Action” transcripts are sometimes created on a rush deadline and accuracy may vary. Please be aware that the authoritative record of the “Research in Action” podcast is the audio.