



The Game Outcomes Project

How Teamwork, Leadership, and Culture
Drive Outcomes


Paul Tozour
Mothership Entertainment

GAME DEVELOPERS CONFERENCE March 14–18, 2016 · Expo: March 16–18, 2016 #GDC16



Hi, my name is Paul Tozour. I'm here to talk about the Game Outcomes Project and what it taught us about the science of creating effective teams, and how teamwork, leadership, and culture drive outcomes.

About Me

- Mothership Entertainment
 @MothershipTeam
- MSE in Tech Management
UPenn/Wharton School
(inspired the study)



Quick introduction: I'm the owner and General Manager of Mothership Entertainment, a small Austin studio building a sci-fi strategy game that we haven't announced yet. I've been in the industry since 1994.

I also earned an MSE in Technology Management from the University of Pennsylvania and the Wharton School of Business, which is only relevant because as I'll discuss later, the inspiration for the Game Outcomes Project came from the Wharton School.

I'm not a consultant, and I'm not trying to sell you anything. I'm a 22-year industry veteran motivated by intellectual curiosity.



Independent industry-academic partnership

- Academia: Dr. Karen Buro (MacEwan U.), Julianna Pillemer (Wharton School)
- Industry: Eric Byron, David Wegbreit, Zhenghua 'Z' Yang, Ben Weber, Lucien Parsons, NDark Teng



I'd like to take the opportunity to introduce the Game Outcomes Project team, who helped design the survey, analyze the results, and write the articles.

We are an independent industry/academic partnership.

Our academic members included Dr. Karen Buro of MacEwan University, who guided our statistical analysis, and Julianna Pillemer, a PhD student at the Wharton School who helped us with a lot of our question design.

Our industry members included Eric Byron, David Wegbreit, Zhenghua 'Z' Yang, Ben Weber, Lucien Parsons, NDark Teng, who translated our articles into Chinese.

I'd also like to thank those of you in the audience who may have participated in the survey; it would not have been possible without you.



1. Survey Design
2. "Non-Cultural" Factors
3. Crunch
4. Culture
5. Non-Correlated Factors
6. Conclusions

GAME DEVELOPERS CONFERENCE March 14-18, 2016 · Expo: March 16-18, 2016 #GDC16



The format of the talk today is divided into the following sections:

1. Survey Design
2. "Non-Cultural" Factors
3. Crunch
4. Culture
5. Non-Correlated Factors
6. Conclusions

Game Outcomes Project Survey

- Asked about most recently completed game
- 116 culture questions
- 4 outcome questions
- 273 completed responses from projects that were neither cancelled nor abandoned
- Resulted in 5 Gamasutra articles

In the fall of 2014, our team designed a survey. We decided to ask game developers a bunch of questions about their most recent game project. This survey included 116 questions about culture, followed by 4 questions about the project's outcome.

And we thought, OK, there have to be at least *some* factors in this list about team culture that are going to correlate with success or failure! And if we can get a couple hundred responses, we can then correlate them and see which of the responses are positively or negatively correlated with different kinds of outcomes.

This was a first-of-its-kind correlational study that correlated a development team's internal culture with their actual results.

We received 273 responses for game projects that had neither been cancelled nor abandoned during development.

Based on that, we did a lot of data analysis and wrote up a

series of 5 articles on Gamasutra.

For those of you who read the Gamasutra articles, I want to assure you that a lot of the data I'm going to present here gives a different analysis from what's in the articles.

Hypothesis:

- Different teams have different cultures.
- These cultural differences *strongly influence* the outcome of a game project.

Is this correct, & if so, which aspects of culture exert the *strongest* influence on a team's odds of success?

Our hypothesis was that different teams have different internal cultures.

We also hypothesized that chance always plays a role, but we felt certain that outcomes have to be *strongly influenced* by their team culture in the aggregate.

And we wanted to know if that was true, and if so, WHICH cultural factors most strongly influence the odds of success, and what can we learn from that.

Game Teams Represented ...

- Rage
- Wasteland 2
- Alien: Isolation
- DC Universe Online
- Sonic Dash
- Assassins' Creed Unity
- Shadow of Mordor
- Donkey Kong Country Returns
- COD: Ghosts
- Dark Dreams Don't Die
- Drive On Moscow
- Second Chance Heroes
- The Dead Linger

... and many more whose identities we don't know!

We had an optional question at the end asking which actual game was represented by each respondent's answer.

Fewer than 10% of our survey respondents answered this question, but based on those who did, we know that our survey represents the following game projects, including many more not listed here.

Part 1: Survey Design



Disclaimers:

1. Not peer-reviewed
2. An *ex post facto* analysis based on a survey
3. Many questions were subjective; subject to biases
4. Correlation != causation

Before we start, a few disclaimers about our study.

1. It wasn't an academic paper, so it wasn't peer-reviewed.
2. It's an *ex post facto* analysis. That is, it's based on a survey asking people about projects they *recently completed*.

In ideal world, we'd have a time machine, and we could re-run the experiment w/ *exact same team* making the *exact same game*, and then measure the difference in outcomes. In the real world, there's no way to run that kind of experiment.

3. It's subject to *cognitive biases*. I might look back on a project with a disastrous ending and be bitter about that, and that might affect my survey response

4. Finally, correlation isn't causation. Our study gave us a ton of correlations we could look at, but we can't technically *prove* that any of these correlations are causal factors by themselves.

I'll discuss correlation and causation a bit later in the talk.

Survey Design

- Randomized order of cultural questions
- Redundant questions with opposing tones
 - "Did the team do X?"
 - "Did the team avoid doing X?"
- 4 outcome questions at end to avoid bias

A few notes on our survey design and the steps we took to minimize bias.

1. We randomized order of the cultural questions to remove any influence that a single ordering might have on the way the questions were answered.

2. We had a lot of redundant questions where we'd ask the same question twice with opposite tones – for example, asking "Did the team do X?," and later on, asking "Did the team avoid doing X?"

We found that this was actually not at all necessary, as there was an almost a perfect inverse correlation between these two questions.

But it was a nice reinforcement that our questions were being answered honestly and accurately.

3. Finally, we asked our 4 questions about outcomes at the very end of the study, because we didn't want to bias anyone about what we were looking for while answering the culture questions.

"Outcomes"



- Return on investment
- Critical acclaim
- Internal satisfaction
- Project delays

Let's talk about what we mean by "outcomes."

Obviously, there's no single definition of what makes a good outcome. Different game projects have different goals and different definitions of success.

If you're making a big AAA game project, you probably care more than anything else about return-on-investment – good reviews are nice but you're spending tens if not hundreds of millions of dollars on your game, and you want to maximize your returns.

Or maybe you're doing a small indie project, and you care more about critical acclaim and building a name for yourself and your studio.

Maybe it's your very first game project, and you'll consider it a success if it just gets done.

Or maybe you're making a political point. Maybe you want to make a game that kills fascists like Woody Guthrie's guitar. And that's fine, so in that case, maybe you care more about your team's internal satisfaction with meeting your project goals.

Our team came up with 4 factors that we felt most developers would agree generally describe positive outcomes.

In other words, most developers would say that a game was successful if it did well on all four of these scales.

- Return on Investment
 - "To the best of your knowledge, what was the game's financial return on investment (ROI)? In other words, what kind of profit or loss did the company developing the game take as a result of publication?" (7-point scale)
- Project Delays
 - "For the game's primary target platform, was the project ever delayed from its original release date, or was it cancelled?" (7-point scale)
- Critical Acclaim
 - "To the best of your knowledge, was the game a critical success?" (7-point scale, laid out specific MetaCritic range)
- Internal Satisfaction / Achievement of Project Goals
 - "Finally, did the game meet its internal goals? In other words, was the team happy with the game they created, and was it at least as good as the game you were trying to make?"

So we asked the following four questions:

"To the best of your knowledge, what was the game's financial return on investment (ROI)? In other words, what kind of profit or loss did the company developing the game take as a result of publication?" (7-point scale)

"For the game's primary target platform, was the project ever delayed from its original release date, or was it cancelled?" (7-point scale)

"To the best of your knowledge, was the game a critical success?" (7-point scale, laid out specific MetaCritic range)

"Finally, did the game meet its internal goals? In

other words, was the team happy with the game they created, and was it at least as good as the game you were trying to make?"

All of these questions were asked on a 6- or 7-point scale.

We then were able to take each question and normalize it to a 0-to-1 scale, with a '0' representing the worst outcome and a '1' indicating the best outcome.

- Return on Investment
- Project Delays
- Critical Acclaim
- Internal Satisfaction / achievement of goals



- Aggregate Outcome
(A score based on adding the 4 factors above)



We felt made sense to combine all four of these into an aggregate outcome score.

Our view was that a truly successful project was one that:

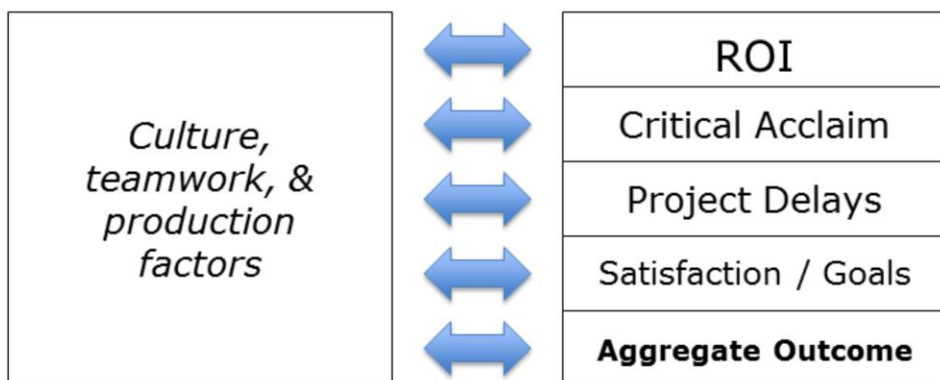
- had a good financial R.O.I.,
- was completed on time,
- received high critical acclaim,
- & that the team felt achieved its goals.

We tried a bunch of different ways of building that aggregate outcome score – we tried approaches based on probability theory (treating the outcomes as confidence values and multiplying them together), and building weighted sums, with a different weighting for each of the outcome factors.

But in the end, we found that simply adding the four individual outcome values together to make an **aggregate outcome score** correlated better with all of the other questions in the survey than anything else we came up with.

So we took those four 0-1 scores, added them together to get an aggregate outcome score between 0 and 4, and then multiplied that by 25 to get a 0-100 score, just like a grade in school.

Correlation and Causation



So we have a whole bunch of cultural factors, as represented by over 110 survey questions.

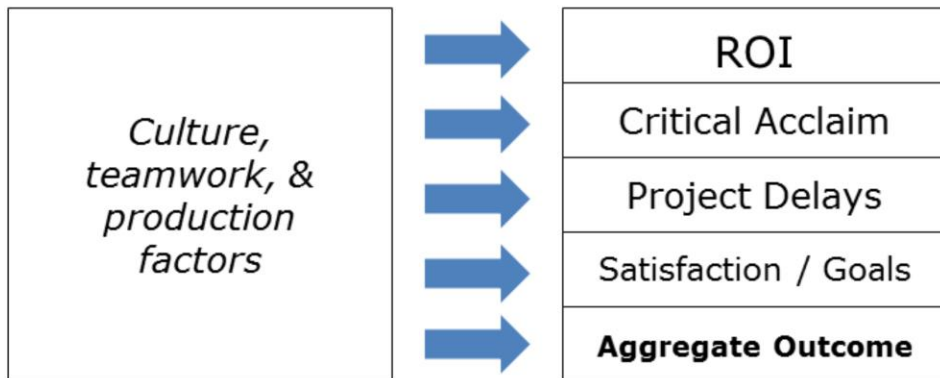
And we can correlate any individual culture question or group of questions with any one of these outcomes, or with the aggregate outcome score

But what does that correlation mean?

For example: maybe we find that teams that used more outsourced labor instead of internal labor influences the aggregate outcome in some way – say, there's a positive or negative correlation – what can we read into that?

Well, because the outcomes happened **AFTER** the game shipped, there's really no way the outcomes could have **CAUSED** the cultural factors.

Correlation and Causation



So if there's any causal relationship at all, it has to be the cultural factors *causing* the outcomes.

Now, sure, it's a survey, and it's possible that somebody's memory of the team culture while taking the survey was influenced by the outcome.

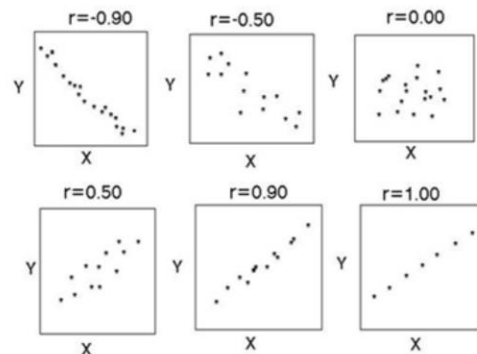
That's possible, and maybe that biased the survey a little bit!

But assuming the bias didn't sway the survey too much, either:

- there is NO causal relationship, or
- t's the cultural factors or some other factors behind them causing the outcomes.

Statistical Analysis

- **Spearman correlation w/** outcome scores for most questions
- **Wilcoxon Rank-Sum Test**
 - ... to detect significant differences between median outcomes in questions with **two mutually exclusive answers**
- **Kruskal-Wallis Test (K/W One-Way ANOVA)**
 - ... to detect significant differences between median outcomes in questions with **more than two mutually exclusive answers**



A quick note on the statistics used in our analysis of the survey results.

Most of you are probably familiar with the concept of a correlation, but just to be sure, here's a quick refresher.

The correlation between two variables is the degree to which those variables show a tendency to vary together.

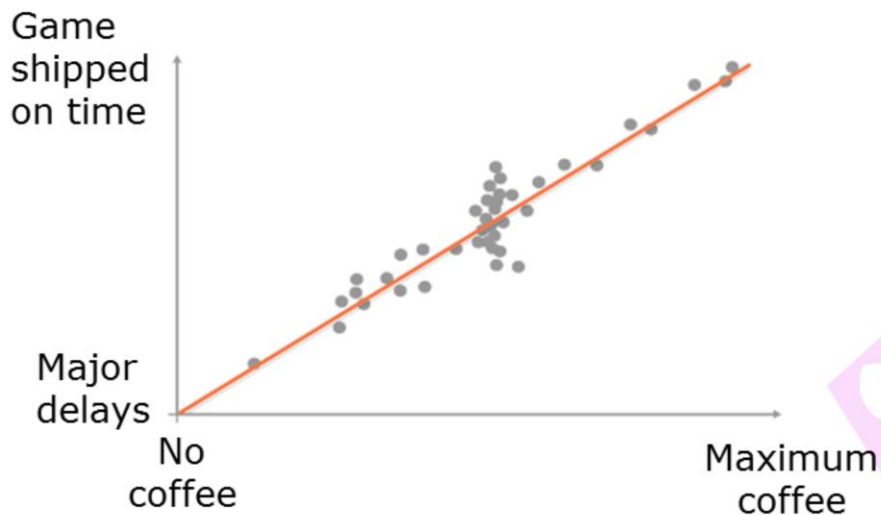
The boxes you see on this slide show an inverse correlation between X and Y (upper left), no correlation (upper right), and various positive correlations (bottom row).

Generally, in the social sciences, a 0.3 correlation is considered pretty darned good.

We used a type of correlations called Spearman correlations for most of the factors in our study because our culture questions used a Likert scale.

We don't really have time to go into the differences between Spearman and Pearson correlations, but I recommend Googling that, and I can assure you that for MOST of the results, the Spearman and Pearson correlation values are very similar, usually within +/- 5% of each other.

We also used a **Wilcoxon Rank-Sum Test** to detect significant differences between median outcomes in questions with two mutually exclusive answers, and a **Kruskal-Wallis Test (One-Way Analysis of Variance)** to detect significant differences between median outcomes in questions with more than two mutually exclusive answers.



Let's assume we have some cultural factor, like teams that drink coffee.

Obviously we didn't ask about coffee-drinking in the survey, but let's roll with it.

And we find that the answers to that question on our survey correlate with higher values on the outcome scale for project delays ("higher" meaning "better," i.e. fewer project delays).

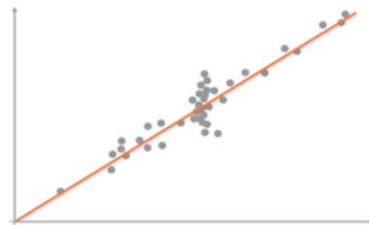
The graph might look something like this. The grey dots form a rising diagonal line, and there's a strong positive statistical correlation between coffee-drinking and teams that shipped their game on time.

What claims can we make about that?

If we have a correlation like this, can we say that drinking more coffee improves your odds of shipping your game on

time?

- 3 possibilities:
 - No causal relationship,
 - A causes B, *OR*
 - Some other factor ("C") causes A and B
- ... But it shows clearly that NOT-A *almost certainly doesn't* cause B!
 - I.e. it disproves the hypothesis that you must AVOID coffee to ship a game on time

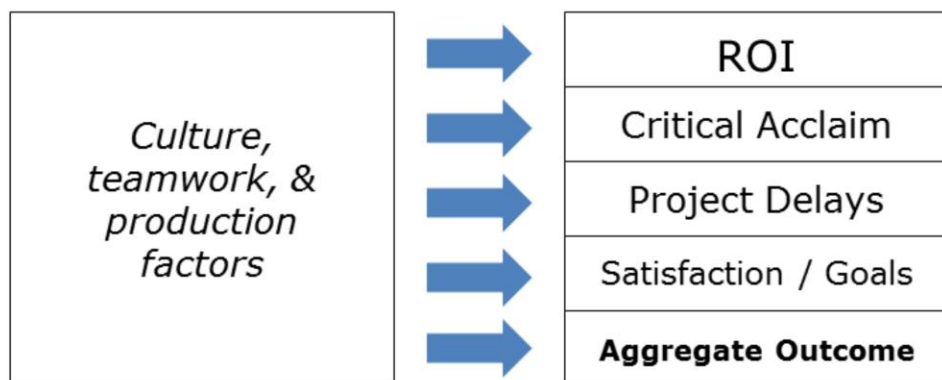


Well, **no**– not exactly. But we *can* say that one of three things must be happening:

- First, there might be NO causal relationship. Depending on the statistical p-value of the correlation, we can rule that out as highly unlikely in some cases. We talk about p-values in the articles, but I don't have time to go into it here; briefly, it's a measure of how likely the results are to have occurred by chance.
- The second possibility is that maybe A really does cause B, and maybe drinking more coffee really **does** help you ship on time!
- Or, 3rd, maybe there's some other, *hidden* causal factor involved. For example, maybe we just hired people who work harder, and maybe hard-working people prefer coffee! That's totally a possibility.

But it *does* let us pretty much rule out the opposite. If someone says, "Hey, drinking coffee actually hurts your chances of shipping a game on time," we can show them this chart, and say, "OK, if that were the case, it should have shown up in this chart. But what we see is the opposite!"

Correlation and Causation



And you can say, OK, for any correlations we find where the statistical p-values are low enough to pretty much rule out coincidence, we can say there could have been some other input C, which accounts for the results. Some *hidden causal factor*.

But we asked over 110 questions over a very broad range of cultural factors, and it's very hard to imagine that there really are too many significant cultural factors that we left out of it!

So we can't 100% rule out the idea that maybe more successful teams just hire better people, and the correlations of all the culture questions we asked were just coincidence, and there's some other magical thing that those "better" people are doing that isn't accounted for in our survey questions.

We can't rule that out entirely, and you're welcome to believe that there are other factors involved.

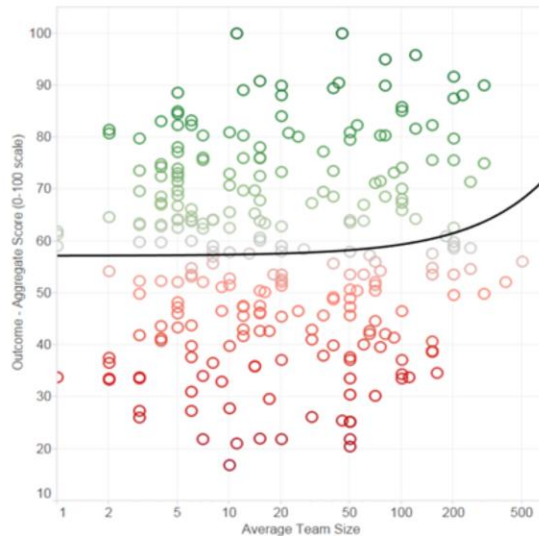
But what I would ask you to think about is this: if you find that there are things that other teams are doing that are *strongly correlated* with success or failure, it might not hurt to learn about them and see how that applies to your own teams. Maybe there's some other factor involved, and maybe not, but if there are simple changes you can make that might improve your odds, why not do that?

Part 2: Non-Cultural Factors

I'm going to start off the discussion of the RESULTS w/ what I'll call "non-cultural factors."

Team Size

- No statistically significant correlation with average or final team size



This is a graph of the average team size plotted against the aggregate outcome.

Each dot represents one team from our study.

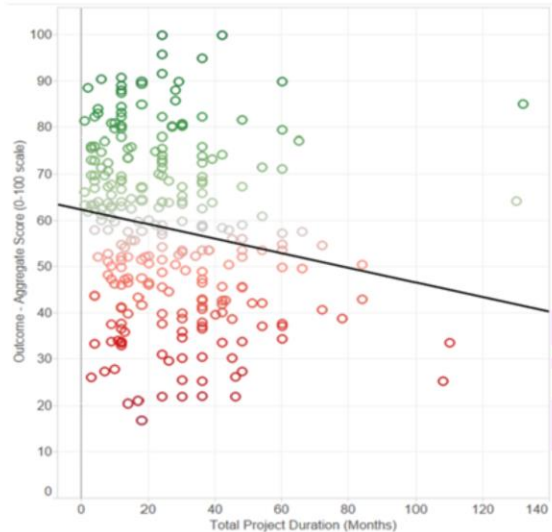
The horizontal axis is team size, based on a logarithmic scale, with larger teams toward the right.

The vertical axis is the outcome, with better outcomes at the top.

As you can see, the trend line goes upward toward the right, indicating that very large teams seem to have slightly better outcomes, although the correlation here is not statistically significant.

Project Duration

- Correlation -0.229
- p-value 0.003

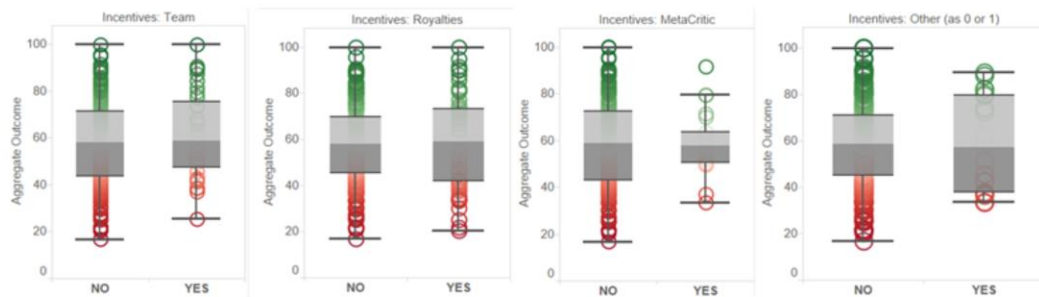


The total project duration had a negative correlation, and *was* statistically significant. You can see that the black line goes downward toward the right.

But that's not really surprising! Projects that are in trouble are naturally going to take longer and experience more project delays.

Incentives

- Most forms of financial incentives had *no statistically significant correlations*



We asked five yes-or-no questions about financial incentives. If you're spending money, or at least offering money, to try to motivate your team, it's interesting to ask whether it actually makes a difference.

And for 4 out of 5 of those, we found that there was absolutely NO significant correlation.

From left to right, we're looking at:

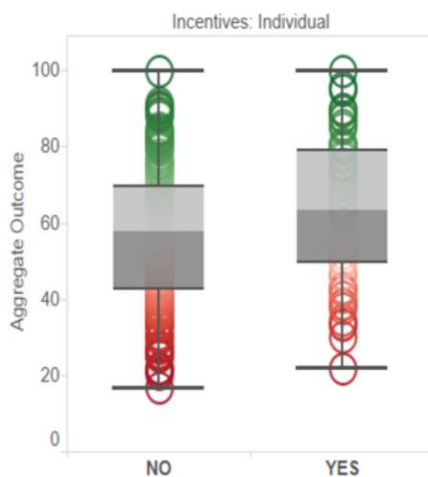
- team-based incentives,
- incentives based on royalties,
- Incentives based on MetaCritic review scores,
- and "other" royalties.

Inside each box, the teams that answered "No" (i.e., no, we did not offer this form of incentive to our team members) are the dots and box-and-whisker plot on the left side, and the ones that answered "Yes" are the dots and box-and-whisker plot on the right side.

We used a Wilcoxon rank-sum test to determine significance.

Incentives

- Individually-tailored
 - Mean outcome:
 - 63.2 for YES (std. dev. 18.6)
 - 56.5 for NO (std. dev. 17.7)
 - p-value 0.017 for the Wilcoxon Rank-Sum test
 - Effect seems to fade for large teams (final team > 50)



The one place where financial incentives *did* make a difference was with individually-tailored incentives. In other words, pay-for-performance plans and things like that.

The mean outcome for teams that *did* use individually-tailored incentives was 63.2, versus 56.5 for teams that did not.

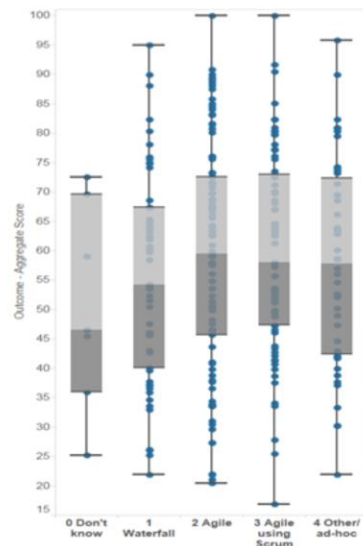
We used a Wilcoxon Rank-Sum test for statistical significance and found a p-value of .017 for this.

We also found that this effect seemed to drop off for large teams, that is, teams of size greater than 50. We're not sure why this would be the case.

However, it *does* suggest that if you're going to offer financial incentives to motivate your team, it's probably a better idea to offer them to *specific individuals* for *specific tasks* within a fixed time frame, along the lines of a Pay-For-Performance (PFP) plan, rather than offering royalties or anything like that.

Production Methodologies

- No correlation with outcomes
- ... or with anything else in the study
- Maybe we put too much emphasis on this?
- Outlier in "Other": Cerny Method



We asked a question about what production methodology a team used, whether it was waterfall, agile, agile using scrum, or "other." We also had an option for "don't know."

And we found that there was essentially NO statistically significant difference. Waterfall seems slightly lower in this image, but it's NOT a statistically significant difference.

This is a very surprising result!

Advocates for agile and Scrum in particular seem to often treat it as a holy grail, and we spend a ton of time talking about methodologies at GDC.

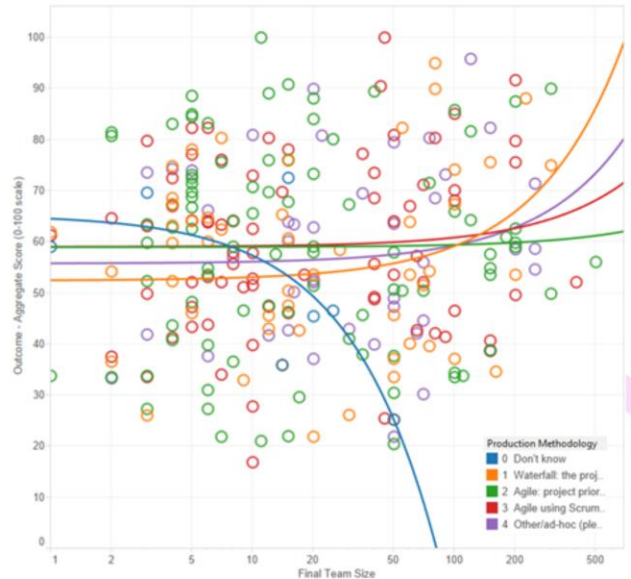
But our results really make you scratch your head. If agile and Scrum are really such holy grails, why don't we see a difference in outcomes?

(Interesting to note: the top-ranking outlier in the "Other"

category reported using the Cerny Method, a methodology unique to game development.)

Production Methodologies

- Agile, Scrum leading slightly for teams of size < 150
- Waterfall leading slightly with teams > 150
- "Don't know" fails for teams > 6



Here's another graph showing the same results in a different way. Each of the colored dots is a single team in our survey.

And we correlate those against the size of the team, which is the horizontal axis, and the aggregate outcome score, which is the vertical axis.

The **orange** dots are teams that used **waterfall planning**.

The **green** dots used **agile**.

The **red** dots used **agile with Scrum**.

The **purple** dots reported using **any "other" methodology not listed here**.

The **blue** dots said **"none / don't know."**

Do you see the pattern?

NO, because there isn't one!

So there's no standout here. Basically, the orange, purple, red,

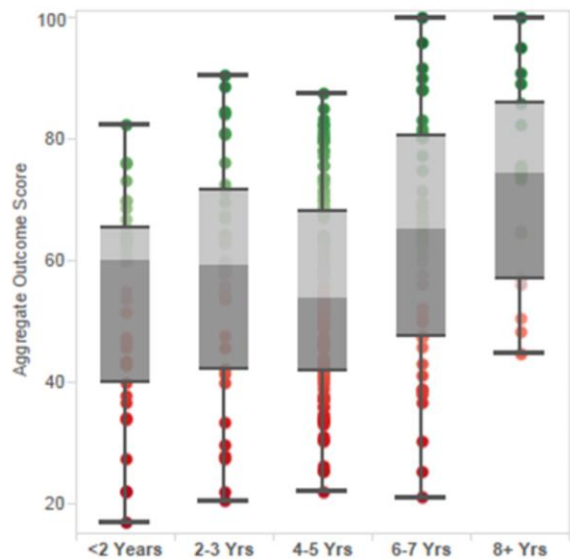
and green lines are *almost the same!* The dots are all over the place but the trend lines are almost identical.

What we *do* see is that “don’t know” response (the blue line) tends to fail for teams larger than 5 or so, which is about what you would expect. Having *no* production methodology doesn’t scale.

And I want to be clear here that I’m *not* saying production is pointless or in any way unimportant. Only that it probably has a lot more to do with *how well* the production is done than whether it uses waterfall, agile, agile with Scrum, or anything else, and how well it’s fitted to your particular game project and your team.

Experience

- Strong, statistically significant correlation
- Experience matters



We also looked at team experience. We asked a question early in the survey about the average level of experience per team member on the team in terms of the number of years, and then measured statistical significance using a Kruskal-Wallis test.

And as you can see, there's a real progression here. Especially as you move from 4-5 years, which is the middle column, to 6-7 years, which is the fourth column, outcomes start to get a lot higher.

Notice how in particular, for teams with 8 or more years of experience, not only is the average much higher, but NONE of the aggregate outcome scores are below 45! The bottom-right corner is almost empty – highly experienced teams seem much less likely to experience serious development failures.

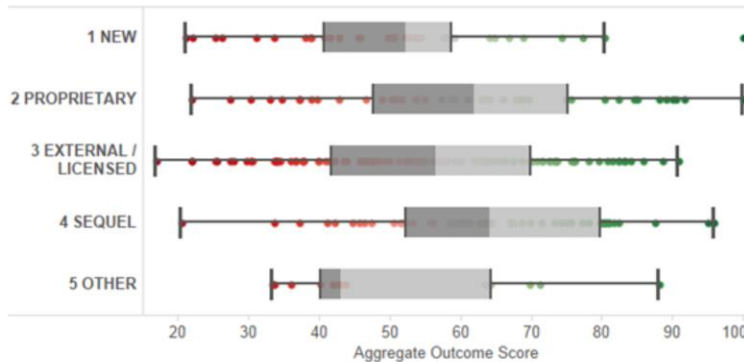
That doesn't necessarily prove that the experience caused the better outcomes. It could simply mean that to a large extent, more experienced developers are attracted to teams that have

historically been more successful, and their experience makes them more likely to be hired by those teams.

Or maybe teams that are more successful tend to hand on to their best members, and their experience gives them time to weed out the least-effective team members.

But either way, experience clearly DOES matter.

Technology Solution



- Only **"Sequel"** correlates with higher outcomes
- **"External / licensed"** correlates with smaller teams, shorter projects (indies)

We asked a question about what technology solution the team used to build its game. From top to bottom, the rows are:

- Top row: a new engine or technology solution created specifically for this game.
- 2nd row: an internal or proprietary engine or technology, such as EA Frostbite.
- 3rd row: an external engine or technology, such as Unity or Unreal or Crytek.
- 4th row: technology or engine from a previous version of the same game or a similar game.
- Bottom row: "other."

And what we found was that *none* of these really had any statistically significant correlation with the project outcome, EXCEPT for 4th column –

– "engine for a sequel based on the previous game's engine."

And that shouldn't be surprising at all, because if you're using

a technology engine from a previous game, that means that:

- the previous game was already successful enough for you to want to make another one;
- so your product and your market have already been validated;
- And you probably already know who your customers are and how to reach them;
- And your engine is a good fit for the game you're making;
- And you probably already have the same team in place that made that game, and they've already mastered that game engine.

So there's a whole host of advantages for projects in this category that can easily explain the statistically-significantly better outcomes of that fourth row.

Part 3: Crunch



Crunch

"Crunch will always exist in studios that strive for quality"

*(extraordinary results
= extraordinary effort
... = extraordinary overtime)*



So let's talk about crunch.

There's an attitude among many developers that crunch is a necessary part of any *good* game development project, and that anybody who doesn't crunch, or who isn't willing to do it, isn't being a team player and doesn't care about quality.

There's a quote from an interview with a certain veteran game developer that sums up this attitude: "**crunch will always exist in studios that strive for quality.**"

And a large percentage of our industry shares this view.

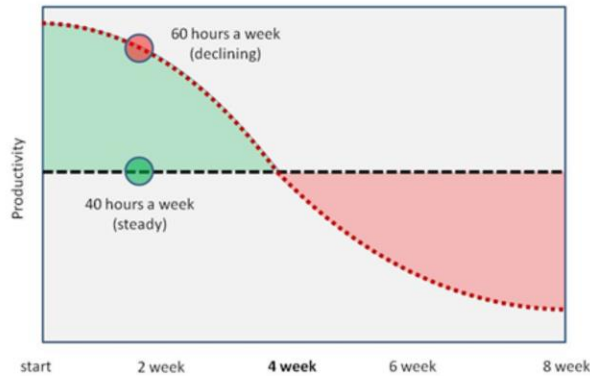
And I want to point out that the *implicit assumption* here is that extraordinary results can only come from extraordinary effort ...

... and that extraordinary effort *necessarily implies extraordinary overtime.*

Now, I think most of us are aware of the data of the harmful effects of extended overtime on *individuals* ...

Crunch

Graphing productivity and overtime



Source:

<http://lunar.lostgarden.com/Rules%20of%20Productivity.pdf>

Also: Mark Layton's "Racing in Reverse" on YouTube

There's **tons** of data clearly illustrating that the abuse of overtime not only leads to lower total productivity, and not only drives a lot of talented people away from the industry, but it leads to a higher risk of relationship failure, mental illness, alcohol abuse, depression, and a host of other problems.

I think many of us are also aware of the recent that pulling all-nighters causes permanent brain damage.

And outside the game industry, it's generally accepted that overtime abuse is a bad idea, and the practice of crunch is relatively rare.

But there's an implicit assumption here, first of all, that crunch *actually improves the quality of a game*.

But how do we know that that's true?

And sure, I know a lot of you are looking at me like I'm crazy, and thinking, "Hey, waitamminute! On my last game we crunched 80 hour weeks for a year and we never would have finished it if we hadn't crunched!"

... And yeah, maybe that's true! And sure, there have been quite a few undeniably *great* games with crunch.

But I also know a lady who chain-smoked and lived to 95. Can we say that because she was chain-smoking until 95, the chain-smoking *caused* her longer lifespan? Or should we look at the *data*, & say, "OK, putting it in perspective, that lady is an *extreme outlier*; it actually takes off 7 years on average?"

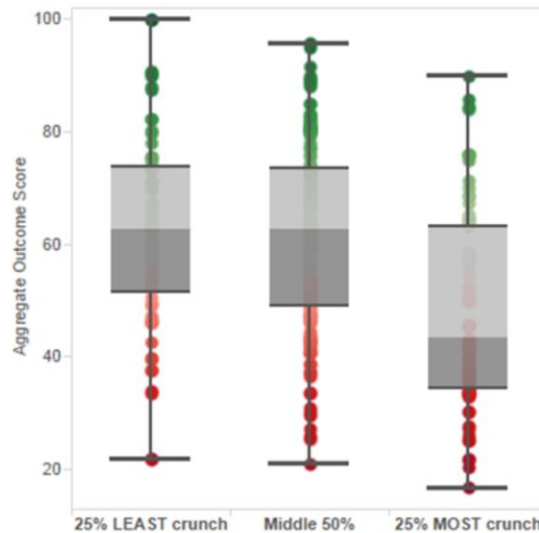
How can we know for sure that that lady wouldn't have lived to 112 if she *hadn't* been a chain-smoker?

... And similarly, if you ship a great game, and you crunched to make it, can you really say "Our game was great because of crunch?" How can you be sure it wouldn't have been *even better* if you hadn't? You don't have the time machine to go back and re-run the experiment.

So I want to ask, is there *really* a solid link in the aggregate between more crunch and better outcomes?

Or are we just *assuming* that crunch works without any real aggregate data to back it up?

Crunch



We asked a total of five questions relating to crunch in our survey. Two of those questions related to the overall amount of crunch that the team engaged in.

The three bars in this graph show our samples grouped into 3 categories.

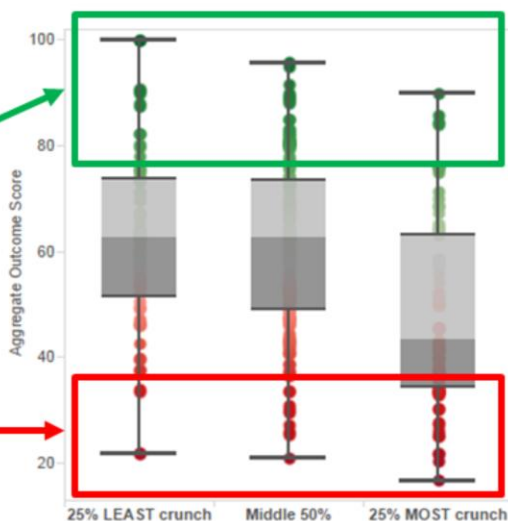
- The bar on the left shows the 25% of teams that engaged in the *least* amount of crunch.
- The bar on the right shows the 25% of teams that engaged in the *greatest* amount of crunch.
- And the bar in the middle is the central 50% of teams, that did a *medium* amount of crunch.

... And the vertical axis is the aggregate outcome score, with the top of the scale representing teams with the most strongly positive outcomes.

Crunch

The games with the most POSITIVE outcomes were in the category with the LEAST crunch

The games with the most NEGATIVE outcomes were in the category with the MOST crunch



So, that makes you wonder. If crunch is a *necessary precondition* for quality, how do you explain the green dots on the top left?

We're looking at a whole lot of successful teams there, and they achieved their success with the *least* amount of crunch.

In fact, the group with the least crunch had the *greatest* number of projects with positive outcomes of all three groups!

I'm sure the green data points inside that box "strived for quality."

And yet they didn't crunch very much, if at all. So how did they manage to achieve success?

And similarly, take a look at the lower right.

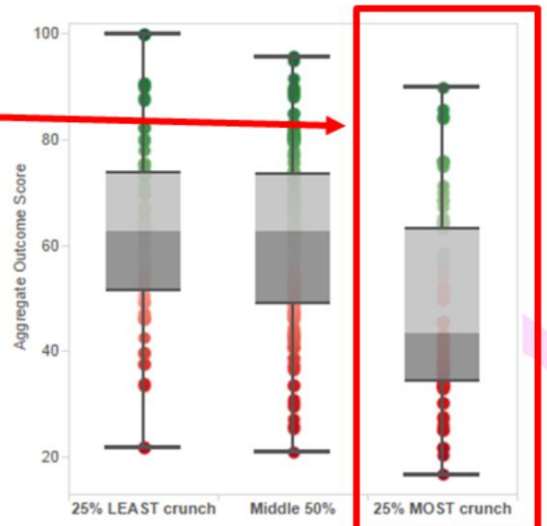
We see a whole lot of red dots toward the bottom, showing teams that experienced very poor outcomes despite doing more crunch than anyone else.

In fact, most of the teams that experienced poor outcomes, used the ***greatest*** amount of crunch.

Crunch

- 25% **MOST** crunch (right):
 - 71 data points, average score 48
 - Middle 50% between 34 and 63

Upper Whisker: 89.95
Upper Quartile: 63.21
Median: 43.40
Lower Quartile: 34.40
Lower Whisker: 16.90



Let's dive in a little deeper.

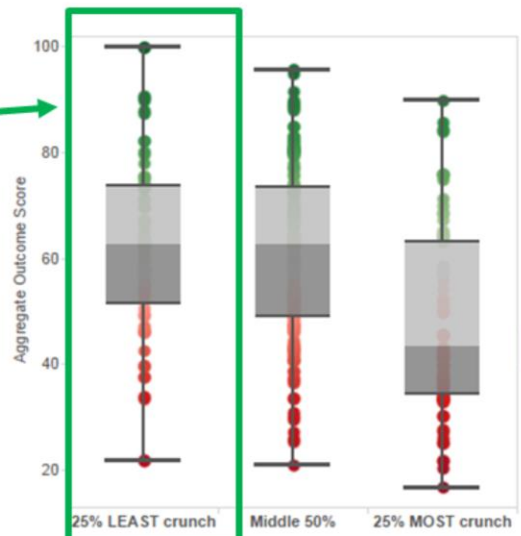
Among the teams that did the *most* crunch, on the right column, there were 71 data points, with an average outcome score of 48.

The middle 50% of projects in this category -- the dots inside the grey boxes in the vertical box-and-whisker plot -- had outcome scores between 34 and 63.

Crunch

- 25% **LEAST** crunch (left):
 - 64 data points, average score 62
 - Middle 50% between 51 and 74

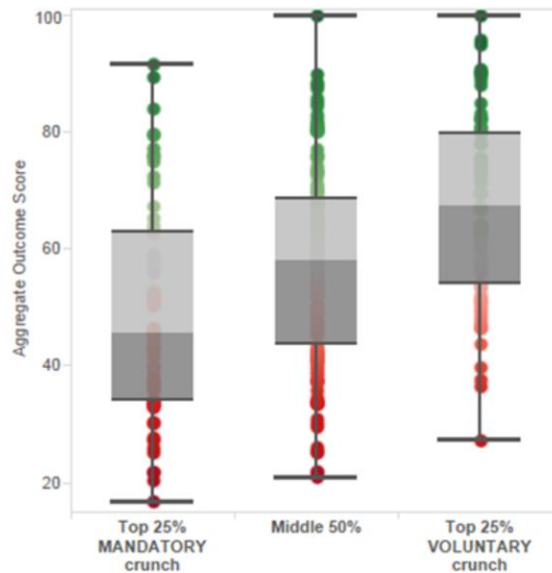
Upper Whisker: 100.00
Upper Quartile: 73.79
Median: 62.57
Lower Quartile: 51.43
Lower Whisker: 21.90



Now let's look at the bar on the left.

These are the 25% of teams that did the *least* amount of crunch. Here, we have 64 data points, with an average outcome score of 62. The middle 50% of the samples had outcome scores between 51 and 74.

Voluntary Crunch



So you may be looking at that and saying,

OK, well, what about when it's *voluntary*?

I mean, hey, look at me – I'm doing overtime because I'm passionate about the product, not because I have a boss holding a gun to my head.

So now let's look solely at voluntary crunch. We had two questions on our survey asking whether any crunch done on the project was voluntary or mandatory, regardless of how *much* crunch actually occurred.

The left column is the *mandatory* crunch group.

The right column is the *voluntary* crunch group.

... and once again, the bar in the middle is the middle 50%.

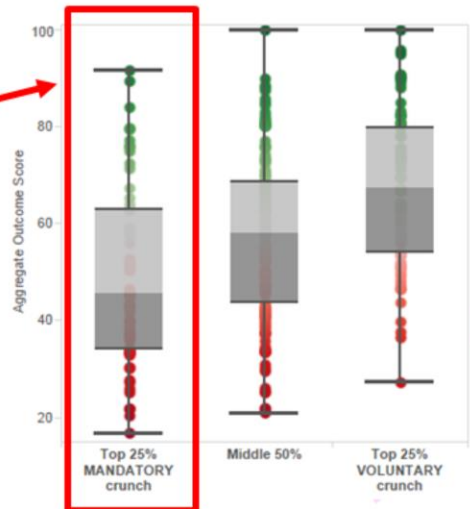
Here, we also see a very strong and statistically significant correlation.

The teams that reported that crunch was mandatory were *far* more likely to report poor outcomes, and the teams that reported that any crunch was voluntary were far more likely to report positive outcomes.

Mandatory Crunch

- 25% most **mandatory**:
 - 60 data points, average outcome score 49
 - 50% between 34 and 63

Upper Whisker: 91.67
Upper Quartile: 63.04
Median: 45.67
Lower Quartile: 34.11
Lower Whisker: 16.90

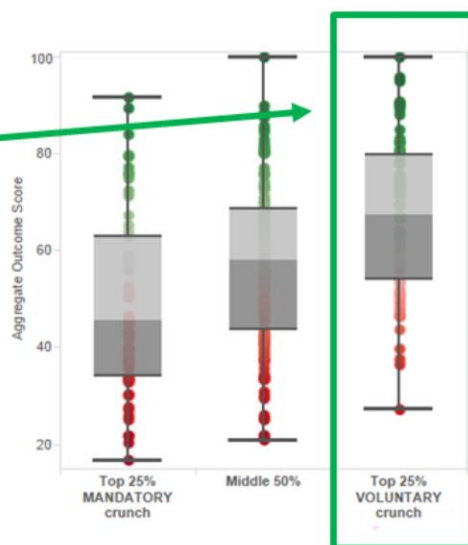


In the “25% most **mandatory**” group, we had 60 data points, with an average outcome score 49. The middle 50% of those were between 34 and 63.

Voluntary Crunch

- 25% most **voluntary**:
 - 71 data points, average outcome score 67
 - 50% between 54 and 80

Upper Whisker: 100.00
Upper Quartile: 79.79
Median: 67.34
Lower Quartile: 54.22
Lower Whisker: 27.34



In the “25% most **voluntary**” group, there were 71 data points, with an average outcome score of 67. The middle 50% of those were between 54 and 80.

Voluntary/Mandatory vs Amount

- Median

	25% Most VOLUNTARY	Middle 50%	25% Most MANDATORY
25% MOST Crunch	<i>Insufficient data</i>	51.17	39.84
Middle 50%	66.79	60.09	52.84
25% LEAST Crunch	70.62	55.30	<i>No data</i>

Here's a graph showing all the data sorted by how much crunch they did, *and* whether that crunch was voluntary or mandatory.

The left column is "voluntary," and the right column is "mandatory."

The top row is the *most* crunch, and the bottom row is the *least* crunch.

Now, let's say you're a newbie coming into the industry for the first time, and you don't have any preconceptions about crunch or whether it's useful or not. In fact, you're so new that maybe you don't even know what crunch is.

Which of these groups would you want to be in?

I don't know about you, but I look at the group in the lower left, with a median score of 70.62, and I find that one really

appealing, not because I'm some sort of lazy person who doesn't want to work weekends, but because they genuinely experienced better outcomes than anybody else.

Now, you may be looking at this and saying, "Wait, hold on a second, Paul! This is all BS. You're picking up the wrong correlation. All you're really picking up is whether the teams were having problems! ... Everybody knows that teams that are experiencing problems are more likely to use crunch to try to solve them. So your analysis is just picking up that underlying correlation!"

Crunch Salvage Hypothesis

"Crunch is more likely to occur on projects that are already in trouble, and these results are just picking up that underlying correlation."

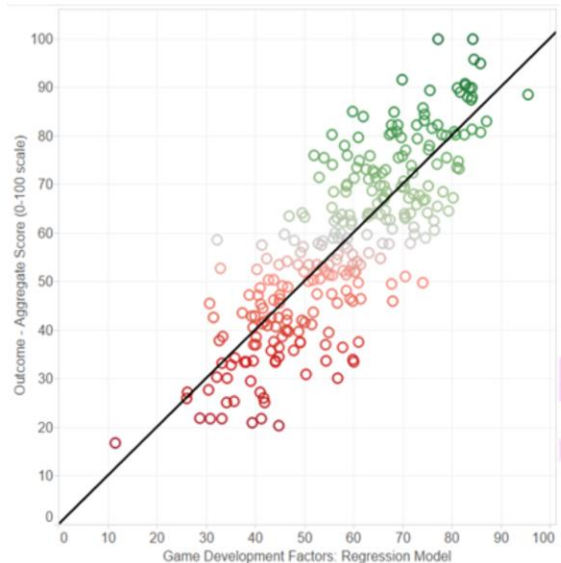
In other words:

"Crunch is more likely to occur on projects that are already in trouble, and these results are just picking up that underlying correlation."

We'll call this the "Crunch Salvage Hypothesis," and we go into this in detail in the fourth article in our Gamasutra article series.

Predictive Model

- Linear regression
- Based on 30 factors
- Correlation 0.82
- p-value < 0.0001



In order to try and tease out whether this was actually skewing our results, we first built predictive linear regression model.

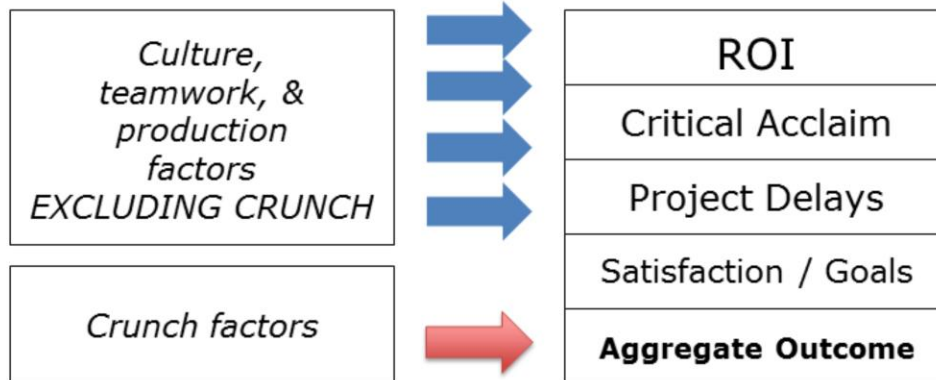
Essentially, we took the top 30 factors in our study, and built a linear regression based on the top 30 factors in our study – mostly cultural factors which I'll discuss later.

This linear regression had an astoundingly good correlation with teams' actual outcomes – it had a correlation of 0.82, and a p-value extremely close to 0.

So the linear regression model could predict the aggregate outcome score for any given team with astonishingly high accuracy.

"Crunch-Free Model"

Correlation 0.81!



How much "lift"?

Then, we took that same linear regression model, and removed all of the inputs related to crunch.

We call this the "crunch-free model."

Removing the crunch-related inputs from the linear regression reduced its accuracy very slightly ... from 0.82 to 0.81.

So this is the box you see in the upper left, taking culture, teamwork, and production factors into account – everything in our study that had a correlation with outcomes, *except* for the factors relating to crunch.

This allowed us to ask, "OK, we know what the crunch-free model predicts that the aggregate outcome score for any given team will be. How much does crunching *actually affect that score*? In other words, what's the difference between the actual score for each team and the score that the crunch-free model predicts, and how does that change with the amount of

crunch?"

And this should have clearly shown us if crunching more actually improved outcomes relative to what they would have otherwise been.

The "Lift" From Added Crunch...

How much crunch affected outcome scores compared to the expected score, as predicted by the "Crunch-Free Model" (same 0-100 scale):

	25% Most VOLUNTARY	Middle 50%	25% Most MANDATORY
25% MOST Crunch	<i>Insufficient data</i>	0.21	-3.91
Middle 50%	-1.69	1.59	4.21
25% LEAST Crunch	3.17	-2.43	<i>No data</i>

So here's the same table we showed 4 slides ago.

Only this time, each box shows the average difference each category's aggregate outcome scores and the scores predicted by the "crunch-free model."

A few things to note here:

- First, these are *tiny* differences on a 0-100 scale. Most of these groups see very small differences from the predicted score.
- And just as before, top right has a *negative* value and lower left has a *positive* value. The box in the top right corner is doing the most crunch and the crunch is the most mandatory, and yet it's experiencing *lower* outcomes than the crunch-free model predicts! And similarly, the box in the lower left – those lazy slackers who aren't doing any crunch – are actually experiencing better outcomes.

This suggests that maybe crunch actually has a *minimal* effect on the outcome!

Another way of putting it is, your studio's *culture* probably already dictates most of the outcome, and crunching isn't going to change that.

The "Lift" From Added Crunch...

How much crunch affected outcome scores compared to the expected score, as predicted by the "Crunch-Free Model" (same 0-100 scale):

	25% Most VOLUNTARY	Middle 50%	25% Most MANDATORY
25% MOST Crunch	<i>Insufficient data</i>	0.21	-3.91
Middle 50%	-1.69	1.59	4.21
25% LEAST Crunch	3.17	-2.43	<i>No data</i>

I want to be clear that we didn't set out with any preconceived notions or trying to prove anything. We just followed where the data led us with an open mind. But the data spoke **very strongly** on this.

Now, you might be saying "Hey, MY team crunched and our game turned out great!" And sure, some projects do a lot of crunch and turn out just fine.

But again, I have to ask: do we crunch because it *works*, or do we do it because we *believe* it works?

A single data point is *meaningless* on its own.

Remember the 95-year-old chain-smoker. There's no way to tell how chain-smoking affects longevity, or how crunch affects project outcomes, unless you do a *thorough statistical analysis* on a large data set.

All of us in this room have different opinions on crunch, and all of us have different experiences. If we go by who's an expert or who's got the most experience, we'll still just end up with a room full of people arguing. The *only* way to settle the debate and go beyond our own very narrow individual experience is with *data*.

Prove Me Wrong!

- Doesn't precisely define "crunch"
- Doesn't nail down point of diminishing returns for OT
- Prove me wrong ... **with data!**

Now, to be fair, there are some potential weaknesses with our findings on crunch.

First, it doesn't precisely define "crunch." How many hours per week for how many weeks? Where do we go from "working a little overtime" to "serious overtime abuse?"

So it doesn't nail down the point of diminishing returns for overtime.

In next version of study later this year, will try to quantify that.

So my challenge to you, if you disagree, is to prove me wrong, but prove me wrong *with data*. Nothing else is going to convince me.

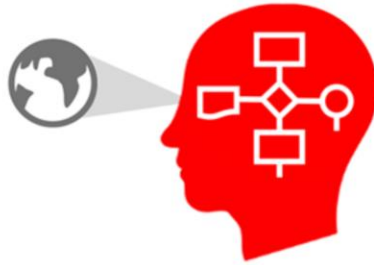
Because anything short of that is just one expert's opinion, and I can easily find another expert who's equally qualified to

take the opposite side of the argument.

Part 4: Culture



Models of Culture



So let's rewind, and talk about what we mean by "culture."

In order to be able to measure culture and then analyze it systematically & quantitatively, we need some sort of conceptual model for it.

And we throw around a lot of words in the industry like "teamwork" and "culture," but those are all very big and very wishy-washy terms.

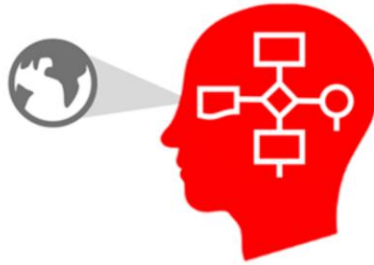
We need to nail down exactly what we mean by "good" teamwork or "bad" teamwork, or how one studio's culture is different from another studio's culture on a very precise level.

And this is IMPORTANT!

You can't debug computer code unless you understand the programming language it's written in.

Similarly, you can't debug your *teams* unless you really understand the *language* of good teams.

Models of Culture



To give you an example of what I'm talking about here: I know some managers who are fans of Freudian psychoanalysis and try to use it as a management tool.

And that is **NOT** the right tool for the job!

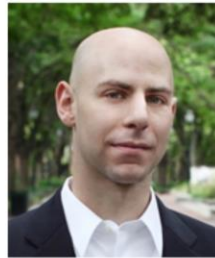
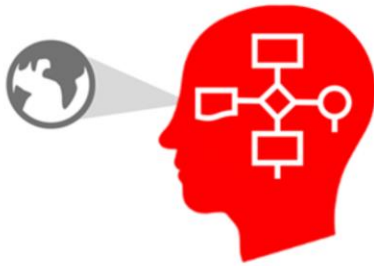
If you're playing armchair psychiatrist, you're *not going to fix anything*.

You're **NOT** going to understand your employees or fix their problems, and you're only going to tick them off and make a dysfunctional team even worse.

So it's important to have the right conceptual model of effective team culture.

Back in 2009-2012, I got an MSE Degree in Technology Management – from the University of Pennsylvania and the Wharton School of Business.

Models of Culture



And as part of that, I studied Organizational Behavior and Design under Adam Grant, the top-rated professor at the Wharton School, and the author of "Give and Take" and "Originals."

And in that class, they taught a *scientifically validated model for team effectiveness* that you could use to analyze and diagnose *any* team in *any* industry.

And I found that fascinating, because I'd never realized there was such a thing before!

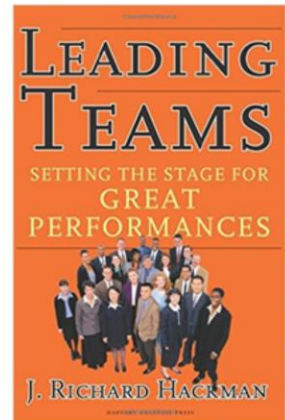
Or that there was such a thing as a "model of team effectiveness," or that it could be scientifically "validated," and as I learned more about it, it put my own experiences in the game industry into an incredible perspective

And I started to look back on my own past experiences and see how so many of the teams I had worked with had different

cultures that correlated *dramatically* with the results they achieved.

Wharton's team effectiveness model

1. Real team
2. Compelling Direction
3. Enabling Structure
4. Supportive Context
5. Expert Coaching



The Wharton team effectiveness model is based on the book "Leading Teams: Setting the Stage for Great Performances" by J Richard Hackman.

So we decided to use Hackman's model as one of the foundations for our survey. Here's Hackman's model in a nutshell:

1. Real team

Clear team boundaries

Team members have clear authority
over own tasks

Team composition based on diverse
skills

Stable membership

2. Compelling Direction

Clear motivating goal

Guides & motivates the team's efforts

3. Enabling structure

Tasks, roles, and responsibilities clearly specified and designed for individual members

Clear definition of who is & who isn't on the team

4. Supportive context

Incentives encouraging the desired behaviors and discouraging the unwanted ones

Tools and affordances to get the job done
– does their hardware & software work?

Psychological safety – can the team speak freely, admit mistakes, and warn of impending problems without fear of blowback?

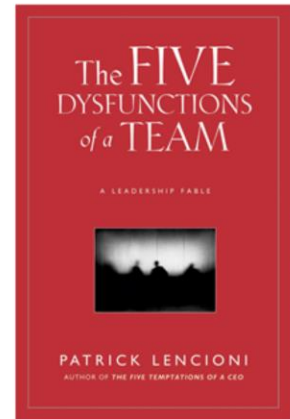
5. Expert coaching

Access to motivators, consultants, & educators outside the team

boundaries who can help raise their game

Five Dysfunctions model

- Absence of Trust
- Fear of Conflict
- Lack of Commitment
- Avoidance of Accountability
- Inattention to Results



In addition to that, we looked at the model described in the book "The Five Dysfunctions of a Team" by Patrick Lencioni/

This book describes a very different model, and it describes how things typically go *wrong* with a team's culture.

Five Dysfunctions model

The FIVE
DYSFUNCTIONS
of a TEAM



Mistrust leads to **fear of conflict**.

Fear of conflict leads to **lack of commitment**.

Lack of commitment leads to **avoidance of accountability**.

Avoidance of accountability leads to a **disconnect from your results**.

Inattention to results leads to games that don't live up to their potential.

I'm going to use Yoda to help me explain this one:

It starts with a **lack of trust** between team members, which leads to a **fear of conflict** because nobody trusts anybody else & no one feels they can speak out safely.

This leads to a **lack of commitment**, because no one feels engaged with the project.

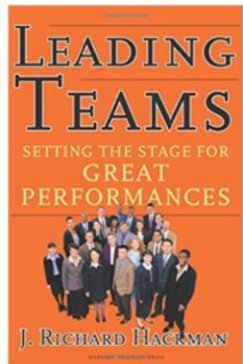
The lack of commitment leads to an **avoidance of accountability**,

and this leads to an **inattention to results** and a disconnect between the team's perception of itself and the results they're actually getting.

And that leads to games that don't live up to their potential.

And that leads to anger, anger leads to hate, hate leads to the dark side, and so on.

Hackman: External environment Lencioni: Internal dynamics



And what's nice about these two models is that they're essentially *orthogonal* to each other.

For the most part, the Wharton School model, that is, the Hackman model describes the external factors, the context and structure that surround a team – how to set up the right environment, motivators, incentives, direction, and so on, and give them access to coaches and motivators.

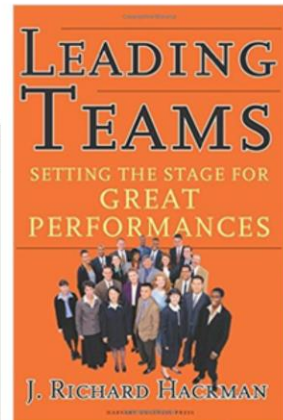
... Whereas the “Five Dysfunctions!” model describes the *internal* dynamics of a team, and how those internal relationships can go awry and what to do to fix them when that happens.

I’m gonna refer to these as the “Hackman” model on the left, and the “Lencioni” model on the right.

Wharton's team effectiveness model

1. Real team: correlation 0.29

		Correlation with:				Category Score
		Project Delays	\$ ROI	MetaCritic	Internal Goals	Aggregate Outcome
1	The organizational structure and membership of the team were clear from the outset of the project.	0.25	0.18	0.18	0.30	0.29
	Most team members had the authority to determine their own tasks on a day-to-day basis.	0.20	0.20	0.16	0.28	0.28
	Most team members were able to determine their own work processes and workflow.	0.16	0.14	0.15	0.25	0.22
	The team composition didn't change during the course of the project (other than growing when needed).	0.27	0.18	0.18	0.24	0.28



So here's what we found.

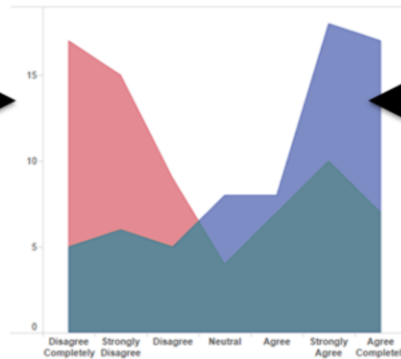
This is one of the charts from our Gamasutra articles.

Now, I'm not going to go through all those numbers here, but if you read the articles, we list the survey questions on the left-hand side, and the various outcome factors such as project delays, ROI, MetaCritic, and internal satisfaction along the top, as well as the aggregate outcome score.

And in green we have the actual correlation values; all of these are positive correlations and are statistically significant.

"The team composition didn't change during the course of development (other than growing when needed)."

Red: bottom 25% of teams by outcome



Blue: top 25% of teams by outcome

But numbers are boring!

And rather than show a whole bunch of correlation values, what I'd like to do is show you the actual responses of the top teams and the bottom teams.

On this slide we have one of those questions from that section, and I've taken just the top 25% of the teams on our survey and the bottom 25% of teams, as ranked by their aggregate outcome score.

So the blue represents the top teams, and the red represents the bottom teams. We've omitted the middle 50% of teams completely, just for these charts in these slides.

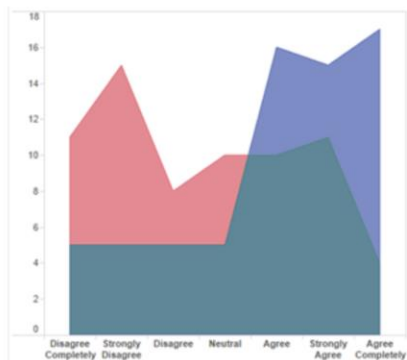
And across the horizontal axis, we have the level of agreement or disagreement with each statement, while the vertical axis of each graph represents the total count of responses in that category for either group.

So look at what happens! The question is: "The team composition didn't change during the course of development other than growing when needed."

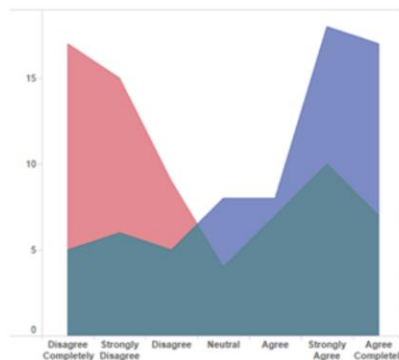
Among the top teams, in blue, 16 people responded "agree completely" and 17 people said "strongly agree." But among the bottom 25% of teams, only 6 people said "agree completely" to this question, and only 10 said "strongly agree."

So the curve is shifted way out to the right. You'll see this with almost all of the questions as we go forward – the top teams agree much more strongly with many of these positive statements about culture.

"The organizational structure and membership of the team were clear from the outset."



"The team composition didn't change during the course of development (other than growing when needed)."



On the left is the survey question, "The organizational structure and membership of the team were clear from the outset." 17 people from the top teams said "agree completely," while only 4 people from the bottom 25% of teams said "agree completely."

And on that same chart, 11 or more people from the bottom teams said either "disagree completely" or "strongly disagree" that the organizational structure and membership were clear, while among the TOP teams, exactly 5 people gave each of those responses.

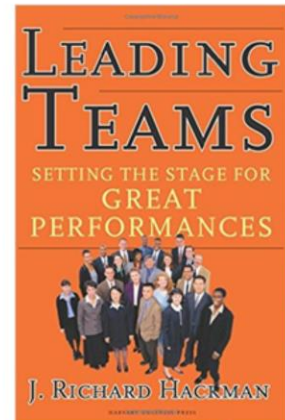
On the right -- "The team composition didn't change during the course of development (other than growing when needed)."

Here, too, we see that the curve is shifted way out to the right for the blue teams and way out to the left for the red teams.

Wharton's team effectiveness model

2. Compelling Direction: correlation 0.56

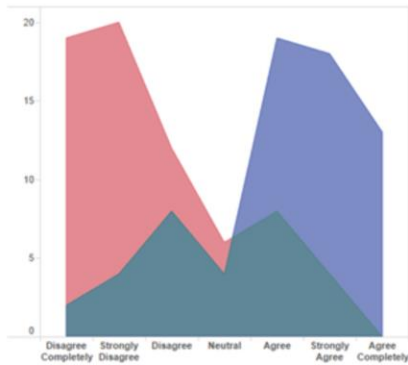
		Correlation with:					Category Score
		Project Delays	\$ ROI	MetaCritic	Internal Goals	Aggregate Outcome	
2	The team believed enthusiastically in the vision for this game.	0.29	0.37	0.41	0.56	0.52	0.56
	The development plan for the game was clear and well-communicated to the team.	0.37	0.39	0.33	0.47	0.51	
	The vision for the final version of the game was clear and well-communicated to the team.	0.38	0.40	0.38	0.56	0.56	



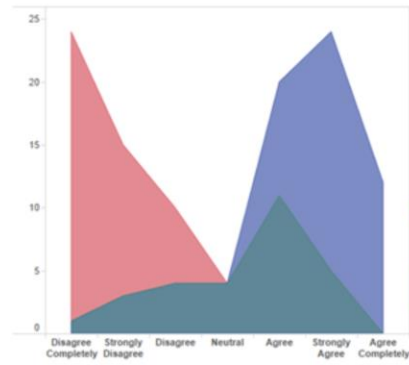
Here are the questions related to the second part of the Hackman model, **Compelling Direction** –

Does the team have a clear motivating goal that helps focus their efforts?

"The development plan for the game was clear and well-communicated to the team."



"The vision for the final version of the game was clear & well-communicated to the team."



On the left:

"The development plan for the game was clear and well-communicated to the team."

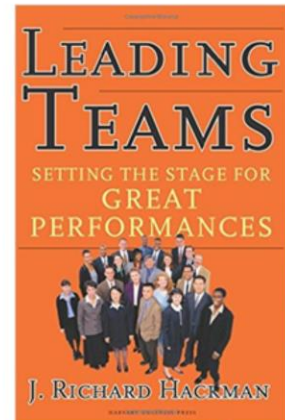
On the right:

"The vision for the final version of the game was clear & well-communicated to the team."

Wharton's team effectiveness model

3. Enabling structure: correlation 0.31

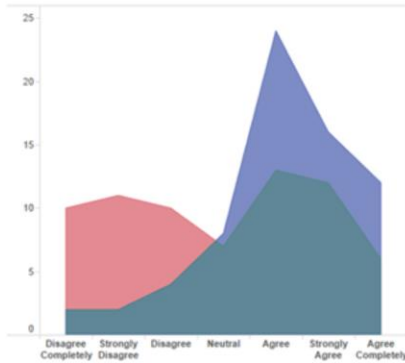
		Correlation with:					Category Score
		Project Delays	\$ ROI	MetaCritic	Internal Goals	Aggregate Outcome	
3	Team members' tasks were well-defined and clearly specified.	0.19	0.16	0.15	0.31	0.26	0.31
	Team members' responsibilities and job roles were carefully matched with their particular skills and abilities.	0.18	0.17	0.23	0.36	0.29	
	My unique skills and talents were valued and utilized while working on this team.	0.18	0.23	0.21	0.38	0.31	



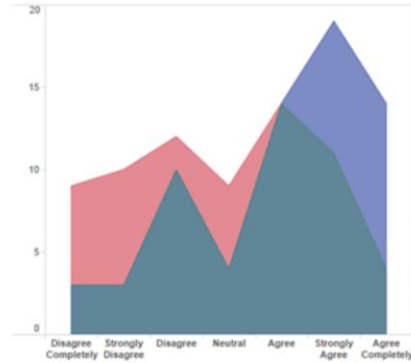
The third part of the Hackman model is **Enabling Structure:**

Are tasks, roles, and responsibilities clearly specified and designed for individual members and properly matched to their individual skills?

"Team members' tasks were well-defined and clearly specified."



"Team members' responsibilities and job roles were carefully matched with their particular skills and abilities."



On the left:

"Team members' tasks were well-defined and clearly specified."

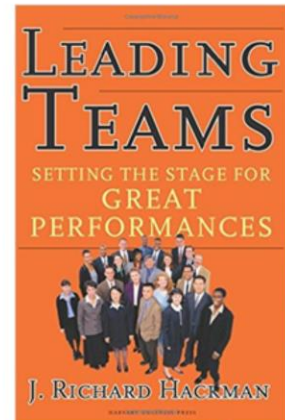
On the right:

"Team members' responsibilities and job roles were carefully matched with their particular skills and abilities."

Wharton's team effectiveness model

4. Supportive context: correlation 0.37

		Correlation with:					Category Score
		Project Delays	\$ ROI	MetaCritic	Internal Goals	Aggregate Outcome	
4	Members of the team were able to bring up problems and tough issues.	0.23	0.28	0.28	0.40	0.37	0.37
	Mistakes were treated as learning opportunities or a chance to improve, not a nail in your coffin.	0.21	0.21	0.24	0.41	0.34	
	Personnel issues within the team (teamwork problems, HR issues) were dealt with professionally and appropriately.	0.30	0.22	0.27	0.35	0.36	

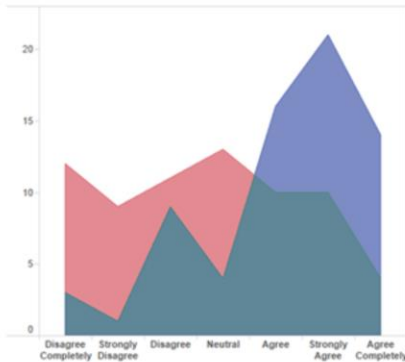


The fourth part of the Hackman model is **Supportive Context:**

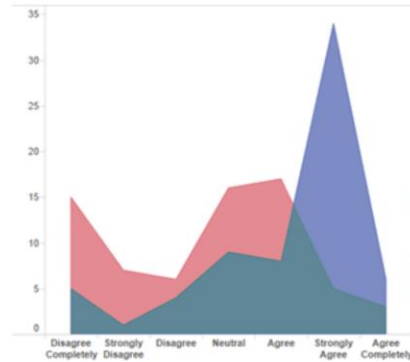
Does the team have the psychological safety they need to speak openly about problems – that is, do they have faith that if they speak up, you'll take appropriate action and they won't be politically damaged by that?

Do they have incentives encouraging the desired behaviors, and the tools and affordances to get their jobs done?

"Mistakes were treated as learning opportunities or a chance to improve, not a nail in your coffin."



"Personnel issues within the team (teamwork problems, HR issues) were dealt with professionally and appropriately."



On the left:

"Mistakes were treated as learning opportunities or a chance to improve, not a nail in your coffin."

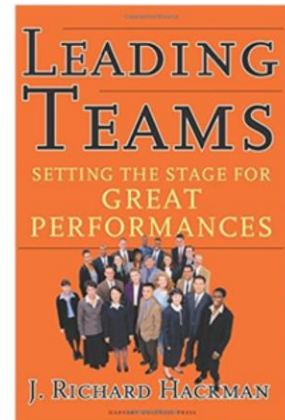
On the right:

"Personnel issues within the team (teamwork problems, HR issues) were dealt with professionally and appropriately."

Wharton's team effectiveness model

5. Expert coaching: correlation 0.27

		Correlation with:					Category Score
		Project Delays	\$ ROI	MetaCritic	Internal Goals	Aggregate Outcome	
5	The team explicitly attempted to develop team members' skills or otherwise assist them in their skill development outside of their normal day-to-day work. We received some form of coaching or guidance to enhance our effort or improve our effectiveness as team members.	0.19	0.20	0.24	0.25	0.27	0.27
		0.15	0.14	0.14	0.22	0.21	

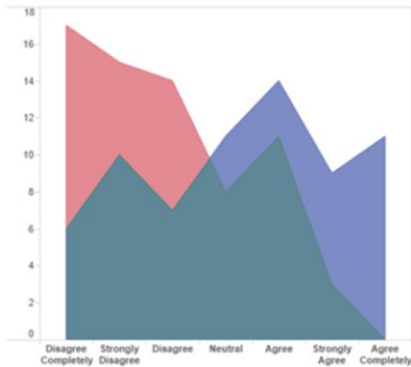


The fifth and final part of the Hackman model is **Expert Coaching**:

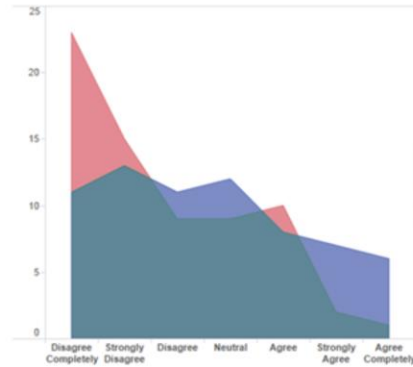
Does the team have access to experts from outside the team boundaries,

including **motivators** to help pull them forward, **consultants** to help guide them, and **educators** to help them improve their skills?

"The team explicitly attempted to develop team members' skills or otherwise assist them in their skill development outside of their normal day-to-day work."



"We received some form of coaching or guidance to enhance our effort or improve our effectiveness as team members."



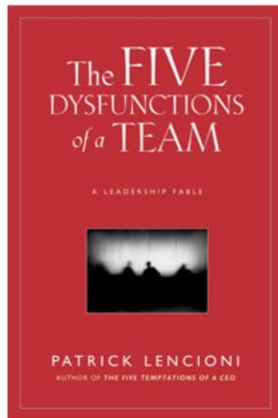
On the left:

"The team explicitly attempted to develop team members' skills or otherwise assist them in their skill development outside of their normal day-to-day work."

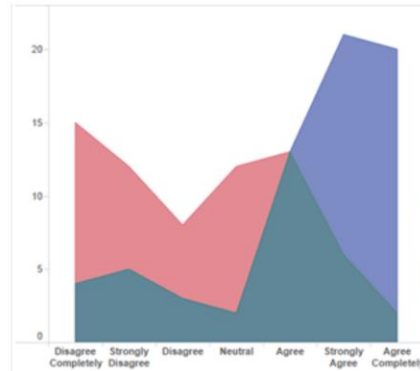
On the right:

"We received some form of coaching or guidance to enhance our effort or improve our effectiveness as team members."

5 Dysfunctions



"It was safe to take a risk on this team and stick your neck out to say something that needed to be said."

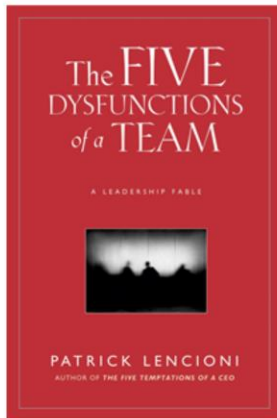


We also asked a bunch of questions related to Lencioni's "Five Dysfunctions" model.

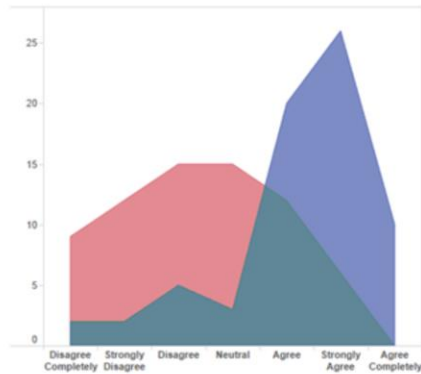
We got very similar correlations here to the strong correlations we got with the Hackman model.

We asked around 20 questions in this category, but in the interest of time, I'm only going to show 3 of those.

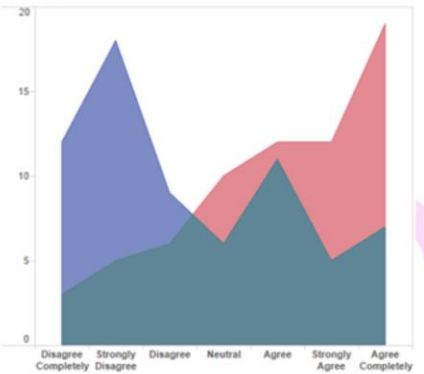
First: "It was safe to take a risk on this team and stick your neck out to say something that needed to be said."



"Most team members bought in to the decisions that were made."



"Some team members put their ego or their own need for recognition above the collective goals of the game project."



In the middle:

"Most team members bought in to the decisions that were made."

On the right:

"Some team members put their ego or their own need for recognition above the collective goals of the game project."

The one on the right is particularly interesting. This question is asked in an inverted tone – about something *bad* that happened on the team. And it's interesting here that the blue and red groups have switched places – the blue teams report much stronger disagreement with this statement, while the red teams report much stronger agreement.

"12: The Elements of Great Managing"

- Mostly redundant with Lencioni / Hackman models
- ... except for:
 - Connection with company's values: 0.39
 - Regular, powerful, insightful feedback: 0.37
 - Peers committed to doing quality work: 0.37
 - My opinion matters: 0.35
 - Recognition and praise: 0.31



We also used a third book, "12: The Elements of Great Managing."

This book is based on a team effectiveness model built by Gallup, which is based on thousands of interviews with teams around the world.

And like the Hackman model used at the Wharton School, this is a scientifically validated model for team effectiveness.

For the most part, this book is the least useful of the three, because most of the factors are already covered by the Hackman model or the Lencioni model.

But there are 5 factors I want to call out that are unique to this book:

- **Connection w/ company's values**

- **Regular, powerful, insightful feedback**
- **Peers are committed to doing quality work**
- **My opinion matters**
- **Recognition & praise**

... And as you can see with the underlined correlation values here, all of those had significant positive correlations with project outcomes.

Additional Game Development Factors ...

- Shared vision for the game: **0.50**
- Resolving disagreements about game's design: **0.45**
- Justifying & communicating design changes: **0.45**
- Celebrating novel ideas even if they don't achieve their intended result: **0.38**
- Respectful relationship between management and team: **0.36**

We also asked over two dozen questions regarding other factors specifically related to game development, which were not tied in with any of the cultural models.

Having a **shared vision for the game** had a correlation of **0.50** with outcomes – this was the strongest single correlation in our study! Clearly, having a shared vision is tremendously important.

Resolving disagreements about game's design also had a very strong correlation of **0.45**

Justifying & communicating design changes had an identical correlation of **0.45**

Celebrating novel ideas even if they don't achieve their intended result was **0.38**, and a

Respectful relationship between management and team had a correlation of **0.36**.

Part 5: Non-Correlated Factors



Here, I want to quickly discuss factors that were *not* correlated with outcomes. As far as we can tell, these factors had no *global* correlations with success or failure.

Non-Correlated Factors

- Team Size
- Production methodology (agile/Scrum/etc)
- Financial incentives (other than individually-tailored ones)
- Cross-functional vs. per-discipline teams
- Game genre (sample sizes too small!)

As previously discussed, team size, production methodology, and financial incentives (excluding individually-tailored ones) had no correlations with outcomes.

We asked about cross-functional teams vs. teams broken down by discipline; there was no global correlation here.

We also asked what game genre each game referred to; there may have been some correlations here but after breaking down 273 responses into over 20 game genres, the sample sizes were too small for any differences to be statistically significant.

Non-Correlated Factors (cont'd)

- Having a friend on the team
- Sharing team members w/ other projects
- Preference for face-to-face communication vs e-mail
- Reliance on temp workers / contractors
- Use of outsourced labor

Having a friend on the team seemed to make no difference.

Sharing team members with other projects, the same.

Having a preference to face-to-face communication vs. e-mail made no difference that we could detect.

Finally, we saw no differences with reliance on temporary workers, contractors, or the use of outsourced labor. For these questions, it's interesting that external labor had no correlation globally.

What the results of these last two questions probably imply is that the real key here is not whether you use external talent, but what external talent you work with, and how carefully and diligently you manage that relationship.

Part 6: Conclusions



Now let's talk about some CONCLUSIONS we can draw from this study.

And this is the part where I get up on my SOAPBOX a bit if I'm not there already.

Conclusions

1. Start demanding *evidence!*
 - Our industry is full of untested assumptions
 - Many practices persist *because we don't question them*

First of all, let's start *demanding evidence!*

The next time someone says waterfall is the devil, and you *must* use agile or Scrum, and your game is ***totally doomed*** if you don't use their production methodology ... ask them for EVIDENCE! Ask them to prove that it *works*.

Our industry is full of *untested assumptions*, and many practices persist ***because we don't question them.***

Similarly, if someone says the ***only*** way to make your game ***truly great*** is to work 80-hour weeks, and your game will be ***crap*** if you don't, ask for evidence of that, too.

Ask for DATA.

Ask for NUMBERS.

Conclusions

2. The real opportunities for improving our odds of better outcomes are mostly

cultural



Second:

The real opportunities for improving our odds of better

outcomes are mostly **cultural**



via Todd Howard,
Executive Producer,
Bethesda Game Studios

Your plan is not as important
as your culture.

In 2012, Bethesda Game Studios Executive Producer Todd Howard gave a keynote speech at DICE, and he showed this slide.

And he said the following, and I'm going to quote him verbatim here because this quote is so appropriate,

"This is one of the big rules that we have, which is, the plan that you have is not as important as your culture.

So you see a lot of game makers will say, 'So here's the big schedule ... here's everything we're gonna do ...'

... you know, if they're really trying, they're gonna run into problems.

And those problems are solved by the **culture** you have on your team."

Every game is a reflection of the team that created it.



I have my own way of putting it, and it's this: I believe every game is a reflection of the team that created it.

If you want to make a better game, you need to build a better team and a better culture on your team.

In other words:

Your GAME comes from your TEAM,

And your TEAM is a product of your CULTURE.

So if you want to make a better game, start by building a better culture.

Conclusions

2. The real opportunities for improving our odds of better outcomes are **cultural**

- Leadership, teamwork, & cultural factors accounted for **>85%** of the difference between the best teams and the rest
- Before you resort to crunch, have you tried:
 - ... debugging your culture?
 - ... debugging your development process?
 - ... making sure the first 40-50 hours / week are actually being used efficiently?

So the real opportunities for improving our odds of better outcomes are cultural.

In our study, leadership, teamwork, and cultural factors accounted for **>85%** of the difference between the best teams and the rest.

And before you resort to crunch, have you tried:

- ... debugging your culture?
- ... debugging your development process?
- ... making sure the first 40-50 hours per week are *actually being used efficiently?* Or are you

trying to scale something that's broken?

Conclusions

3. There is an established **science** to building effective teams
 - Our findings mirror those of management science
 - Be aware of the limitations of your own experience and instincts
 - Our industry is not a unique and special snowflake. We are still humans working in organizations, and management research is equally relevant for us!

Conclusion #3:

There is an established **science** to building effective teams

LET'S START USING IT!

- Our findings mirror those of management science exactly.
- Be aware of the limitations of your own experience and instincts. A lot of us have worked on a lot of games but almost none of us have actually analyzed our

experiences rigorously or worked on enough games to represent a statistically significant sample size.

- Recognize that our industry is not a unique and special snowflake! We are still **humans** working in **organizations**, and there's a huge treasure trove of invaluable management research waiting out there that's equally relevant for us!

4. **Design** your culture intentionally

- Don't just let it "happen"
- Write/update your mission/values statement
 - It matters!
- Hire people who will embrace your values.
- It's not enough to "hire great people and get out of the way."
- Your job as a leader is to **proactively manage risk**

Conclusion #4:

If you want to get the most out of your culture, you have to **design** it. It's not going to happen by itself!

- Don't just let it "happen"
- Write/update your mission/values statement – It matters!
- Hire people who will embrace your values.
- It's not enough to "hire great people and get out of the way." If you really believe this, and that's how you work as a leader in the game industry, you're missing about 95% of your job.
- Your job as a leader is to **proactively manage risk**

Designing your culture intentionally

means it's your job to ...



- Incentivize the behaviors that **align** with your values, dis-incentivize those that **do not**!
- Actively **shape** your team's internal discourse to promote healthy feedback & creative conflict in the creative process and minimize *politics* & *interpersonal* conflict.
- Keep the millions of **small** problems & values misalignments from turning into **big** ones

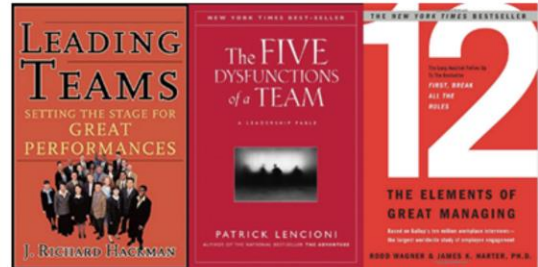
Designing your culture intentionally

means it's your job to ...

- Incentivize the behaviors that align with your values, dis-incentivize those that do not!
- Actively **shape** your team's internal discourse to promote healthy feedback and creative conflict in the creative process and minimize *politics* and *interpersonal conflict*.
- Keep the millions of **small** problems & values misalignments from turning into **big** ones.

Contact Info & Future Projects

- Gamasutra articles: <http://ubm.io/16aqzJv>
- Questions/feedback to  **@GameOutcomes**
- New survey in late 2016
- Leader Interviews:
contact **Eric Byron** at ebyron15@gmail.com
- Me: paul.tozour@gmail.com  **@MothershipTeam**



Thank you for watching!

Our Gamasutra articles are available at <http://ubm.io/16aqzJv>

Please follow @GameOutcomes on Twitter for updates, announcements, occasional additional data analysis, and to ask any questions you may have.

We'll be doing a new version of the survey later this year.

Team member Eric Byron is doing a set of leader interviews related to the Game Outcomes project; I'd encourage you to contact him at the e-mail address listed (ebyron15@gmail.com) if you're a game industry leader interested in participating.

Finally, my contact info is available at the bottom.