Testing

Learnings from MA (and brainstorming):

* Static tiles (obstacles) used very often
* Movable obstacles rather rare, especially in swapping games
* In collapse levels, a lot of space taken by obstacles (~25%)
* Usual board size 7-9 wide and 9-12 high so we can stay with 8x10 as average
* In collapse levels, large variation of matchable space => could test collapse with obstacles and vary obstacle count to see effects
* Generally, varying number of obstacles present seems to be a good way to test scenarios
* Late game often combines obstacles with removable obstacles as a level goal
* Toy blast, a collapse game, uses overlays (blocked pieces and shielded pieces) a lot
* Spawners (slime blocks) are used a lot by both styles in the later levels
* Collapse levels usually have lower count of special tiles
* Normally just one level objective
* General numbers for special pieces: spawners (slimes) 1-5, obstacles, breakable pieces and overlays 5-30, special tiles (time bombs) 1-10
* Will set test-number for each piece as follows: spawners 1/3/5, obstacles etc. 5/10/20, special tiles 1/5/10

Scenarios to test:

Scenarios were chosen based on own research on real levels in candy crush saga, bejeweled, and toy blast, and through consideration of the information in the MA.

Swapping:

* Score without anything else (bombs active) x
* slime + blocked pieces (jellyfish goal) x
* slime + blocked pieces + removable obstacle x
* time bomb + removable obstacle x
* locked pieces + removable obstacle x
* time bomb + locked pieces + removable obstacle x
* shielded pieces + removable obstacles x
* shielded pieces (shielded pieces jellyfish) x
* removable obstacles x
* obstacles (score) x
* removable obstacles + obstacles x
* movable obstacles (score) x

Collapse:

* score without anything else (bombs active) x
* full clear without special tiles x
* removable obstacles x
* slime + blocked pieces x
* slime + blocked pieces + removable obstacle x

**Start of testing**

Collapse clear: consistent in time and move count, rather simple, most thought goes into collapsing the right pieces to create more fields of colour

Collapse score: pretty braindead, and bombs are very useful as you don’t have to match a bomb to make it explode, you can just click on it

Collapse removable: more interesting and some more thinking involved, low obstacle count way too easy, medium and high count seems to not have too much of an effect

Collapse slime blocked: blocked pieces were jellyfish. Fun, more challenging than scenarios before. Difficulty pretty consistent between medium and high piece counts, though low piece count significantly easier

Collapse slime blocked removable: Surprisingly easier than only slime and blocked on every level.

Swapping score: move count and score varies a lot more than with collapse. Most likely because creating matches is more complex with this control type

Swapping obstacles: More interesting than only score. Still not hard, but adds a lot more thinking and is thus more engaging. Becomes highly dependent on RNG with high obstacle count

Swapping moving obstacle: also RNG-dependent, though not as much as static obstacles at high counts. Also causes very interesting board states due to the obstacles moving and being affected by gravity. Feels trickier than static obstacles. At lower counts pretty much feels like normal score though

Swapping removable: Gets very difficult on higher obstacle count. Engaging to play though. Can imagine this getting easier and more engaging with more mechanics added

Swapping removable + obs: More difficult than just removable, though depends on where exactly the obstacles are placed.

Swapping shielded: used 1/5/10 pieces. Very difficult, and dependent on RNG because of the dependence on bombs in the right positions. Strategic swapping required. On highest shielded count, closest to victory was 3 pieces left

Swapping shielded removable: literally almost impossible without a lot of luck on high piece counts

Swapping blocked removable: probably harder even than shielded pieces, but felt more manageable somehow

Swapping blocked removable time: Time bomb had 10 turns timer. Really difficult as well. Blocked pieces in swap is just a very difficult goal, and removing the bombs gets very difficult in high piece count.

Swapping slime blocked: Turns out blocked pieces are just really difficult to remove without gadgets. Very difficult, and slimes are also more of a threat in swap.

Swapping slime blocked removable: easier than just blocked, maybe because of lower number of blocked pieces.

Swapping time removable: Difficult, but not too much, so was actually pretty fun.

**How to analyse data brainstorming**

See how difficulty changes between different tile counts

* How to do that?

Obtain a difficulty rating for each tile count, then look how difficulty mounts

* Reduce the testing data to three entries, with each one being the average of each of the three testing scenarios

In the reduced scenarios, instead of success being true or false, take success rate in %.

* Should total time be included in average if time is failure? And how to interpret total time and time between moves, as it can vary based on level objectives and not necessarily be connected to difficulty?

I think total time should be included in the average of the scenarios. The general workflow should be:

* Take average of testcases for each scenario
* Out of this average, calculate a “difficulty rating”
* See how the difficulty changes between scenarios
* (Somehow come up with a total difficulty rating to be able to rank the testcases independently of tile counts)

Maybe just take difficulty rating of the middle case as total difficulty? Or maybe take average? Could ask Sebastian about it.

Problem with difficulty: high moves don’t mean difficult, because in score levels you just make as many moves as possible. I think a high time between moves is more indicative of difficulty, along with the success rate in %. If a level was successful and not score, a higher total time also indicates higher difficulty imo. I have no idea about score

I think the best way to go forward is:

* Do the average taking as said above
* Determine difficulty somehow, first just looking at the raw values and later with a function maybe
* See if the assigned difficulty aligns reasonably well with the subjective experiences from the testing protocol

Afterwards I can still think of other ways to determine difficulty, but I think this is fine as it is also what I said in my expose, just look at the values and higher = difficult.

When I have the difficulty assessment, I can gather some other statistics as well and show them if I want to.

But don’t forget to see if there are any interesting scenarios to test that arise from evaluating the data!!

Probably something with blocked pieces and swapping, maybe with the paint bomb as a way to more easily get matches would be interesting. Let’s say max 3 more testcases, then we would be at 20

Difficulty assessment: I think it should just include total moves, time between moves, and success rate. Maybe total time for successful attempts. Score is not really indicative of difficulty.

For score levels, difficulty assessment might need to be done differently.

* Higher time, time between moves and total moves indicate higher difficulty. Lower success rate indicates higher difficulty.
* First naïve function: d = 1/success \* (time + movetime + moves)
* When success is 0, doesn’t work
* Just add time, move time and moves, and then see what the general variation is. Based on that, make table to add score based on success rate. Lets try that

Still need other way to calculate difficulty in score levels

**Evaluation Protocol**

Calculated average of total time, time between moves, total moves, and score and the success rate for each scenario and each of the 3 tile frequencies separately.

Put the collected averages in a shared excel spreadsheet.

Calcluated difficulty of each average with the following formula:

D = 100 + (time + movetime + moves) – success \* 100

For levels with time bombs, the game time was counted as 120s if the game was lost due to a time bomb exploding.

Wrote difficulties into table and made diagrams. Interesting findings: